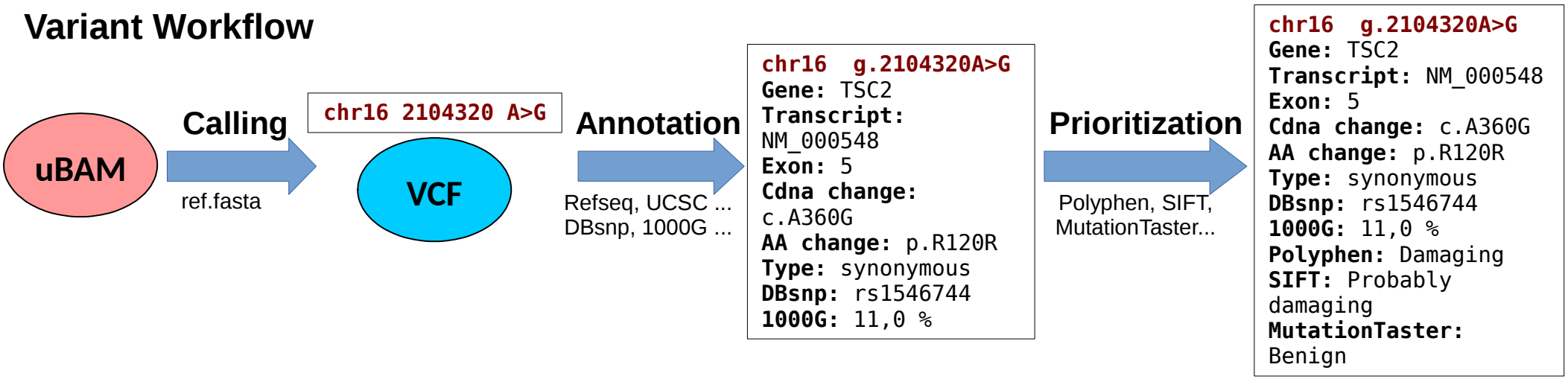
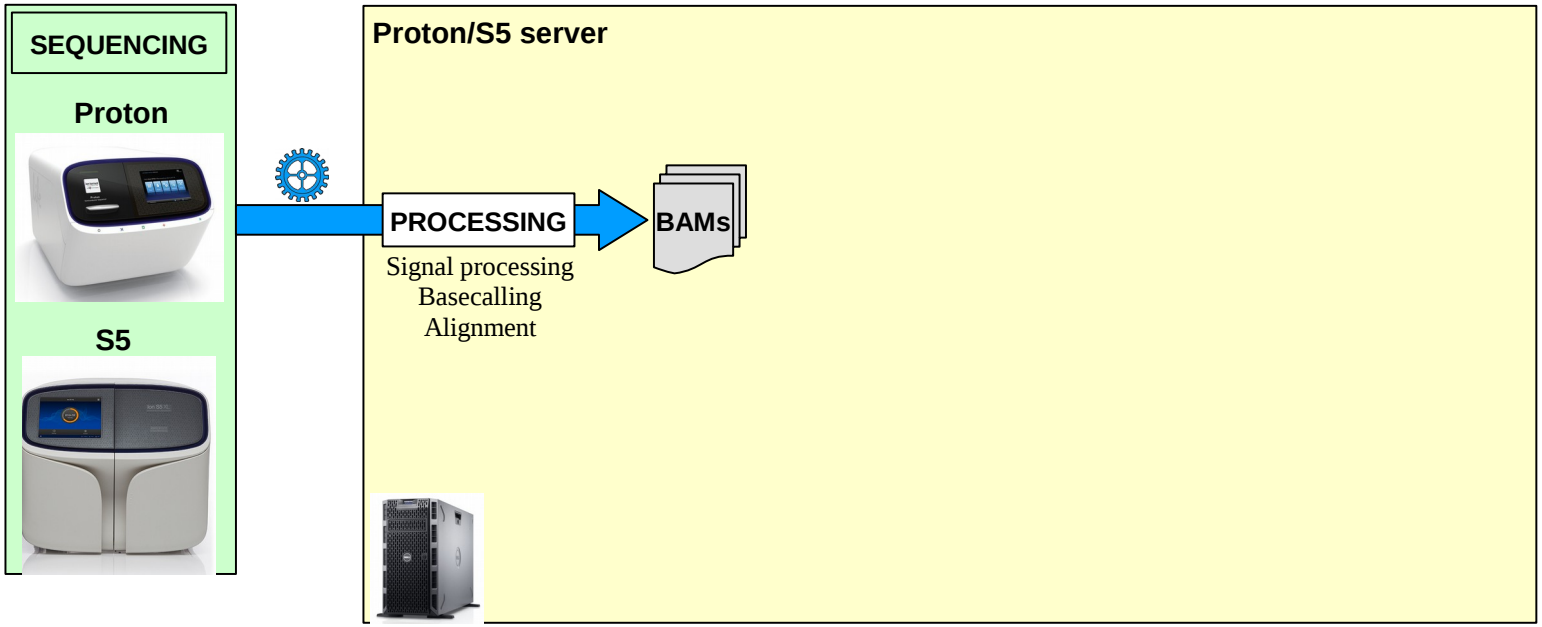
Pipeline Diagnostique NGS

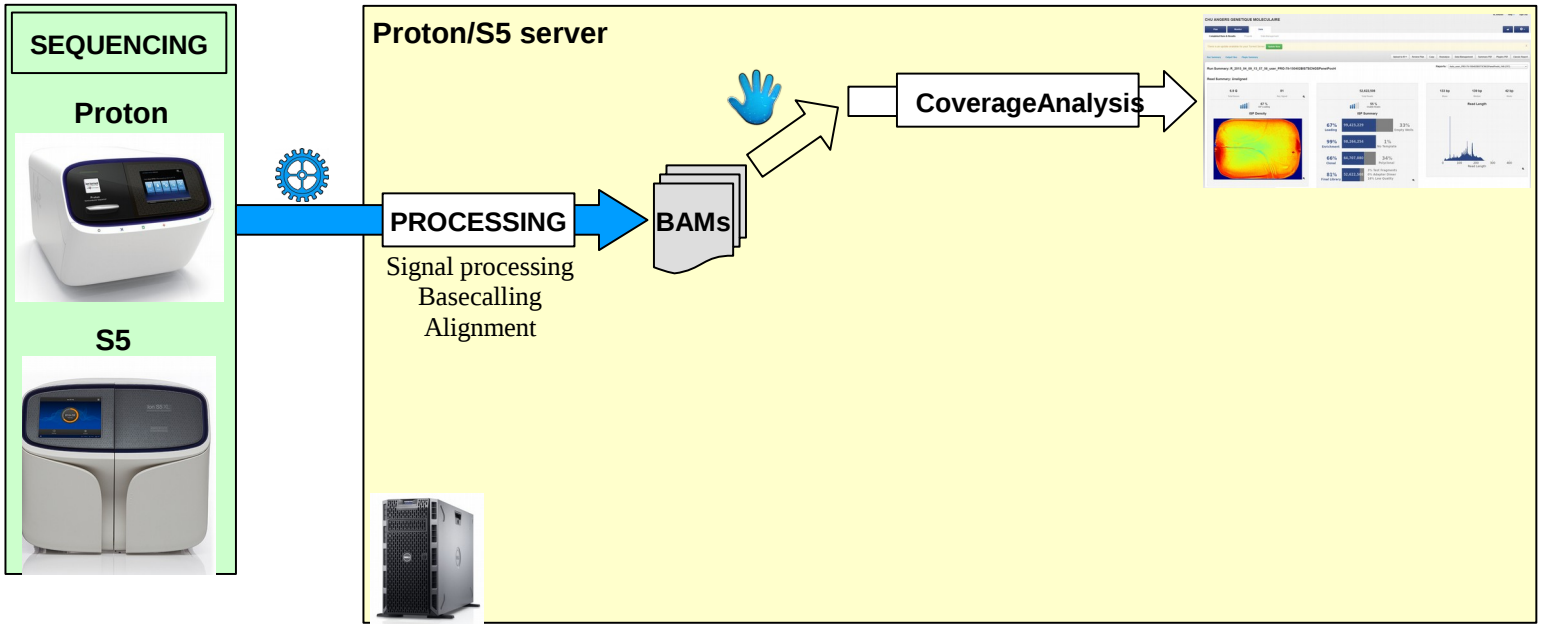
Variant Workflow



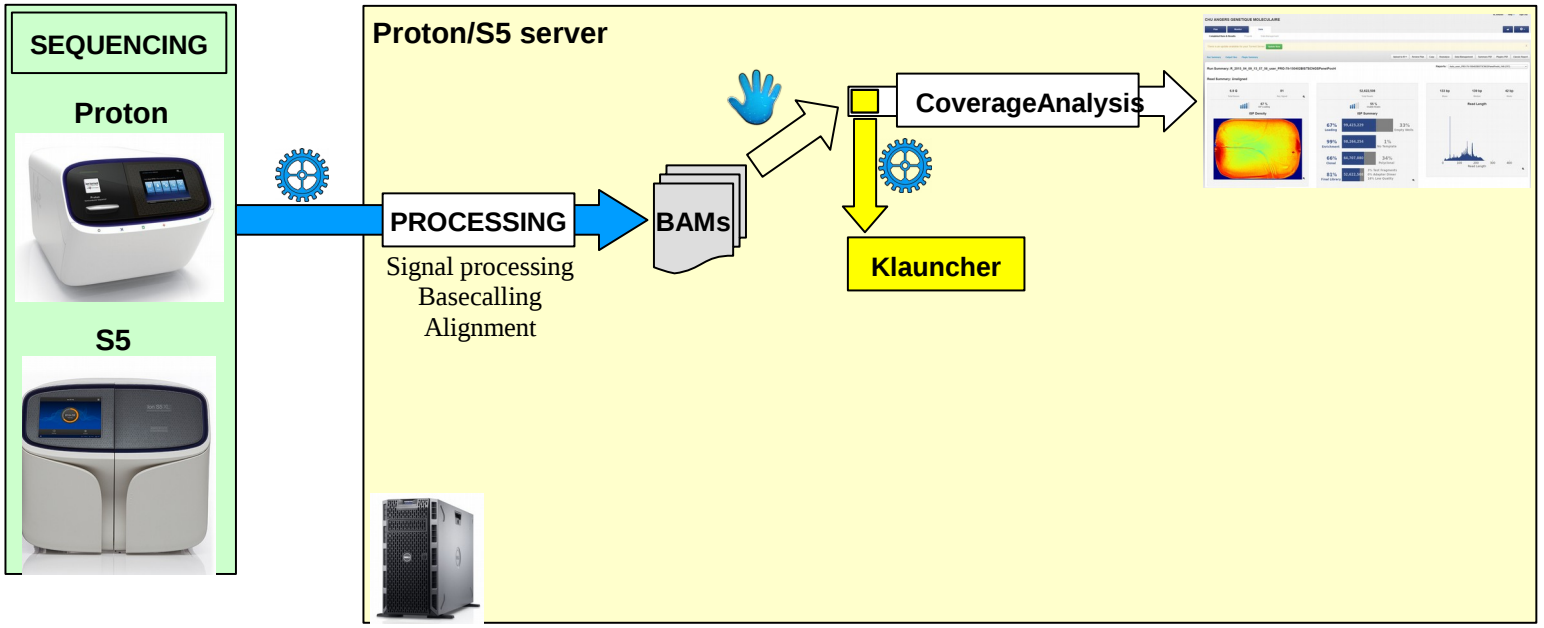
Diagnostique NGS – Pipeline Maison



Diagnostique NGS – Pipeline Maison

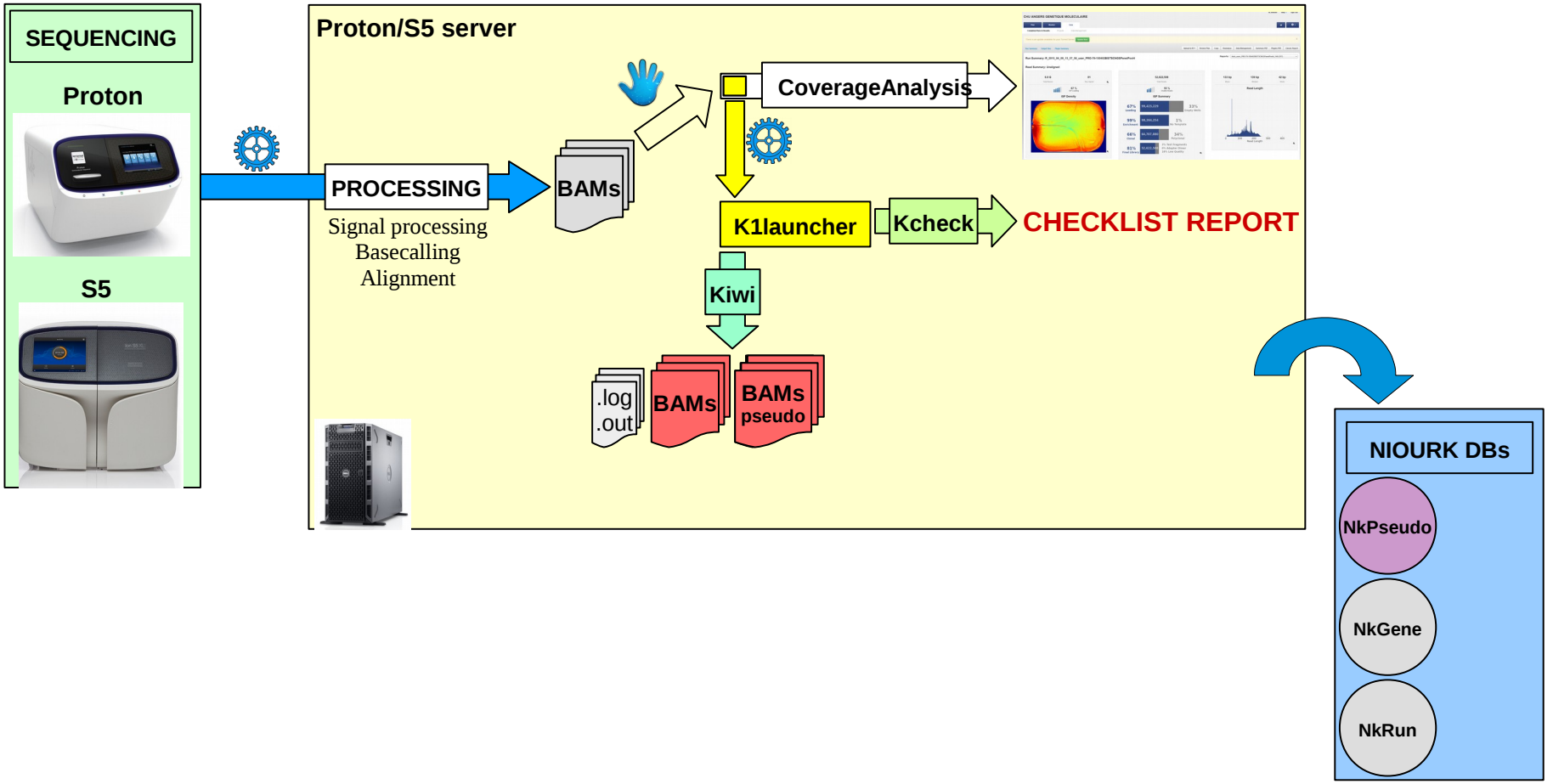


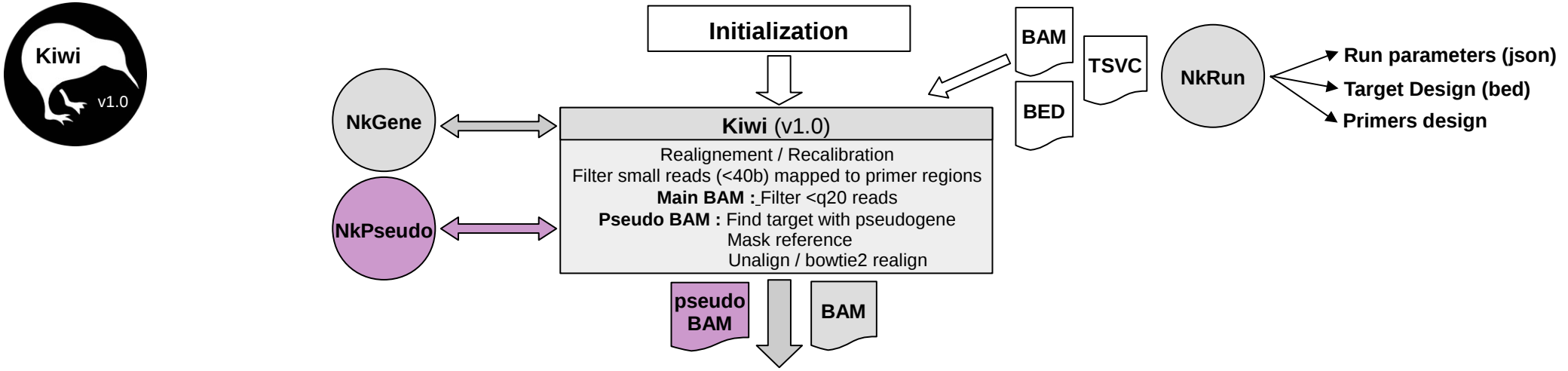
Diagnostique NGS – Pipeline Maison



```
 _ ___                                  _
| |/ / | __ _ _   _ _ __   ___| |__   ___ _ __
|   /| |/ _` | | | | '_ \ / __| '_ \ / _ \ '__|
| . \| | (_| | |_| | | | | (__| | | |  __/ |
|_|\_\_|\___|\___|_| |_|\___|_| |_|\___|_|  [v1.0]  (Mon Feb 29 10:20:12 CET 2016)


Processing run: Auto_user_Proton-187-160118TSCCUSTOMPANEL4POOLHIQ_287_451


  < Initialization >

  < Processing samples >

    IonXpress_015
        Kiwi      TVC      TVC pseudo      Niourk
    IonXpress_017
        Kiwi      TVC      TVC pseudo      Niourk
    IonXpress_018
        Kiwi      TVC      TVC pseudo      Niourk
    IonXpress_019
        Kiwi      TVC      TVC pseudo      Niourk
    IonXpress_022
        Kiwi      TVC      TVC pseudo      Niourk
    IonXpress_023
        Kiwi      TVC      TVC pseudo      Niourk
    IonXpress_024
        Kiwi      TVC      TVC pseudo      Niourk
    IonXpress_025
        Kiwi      TVC      TVC pseudo      Niourk
    IonXpress_026
        Kiwi      TVC      TVC pseudo      Niourk
    IonXpress_027
        Kiwi      TVC      TVC pseudo      Niourk
    IonXpress_028
        Kiwi      TVC      TVC pseudo      Niourk
    IonXpress_029
        Kiwi      TVC      TVC pseudo      Niourk

  < Niourk add frequency >

  < Niourk Excel >

  < BlanKet >

  < Cleaning & Copy >


Elapsed time: 6h:36m:52s
```

Diagnostique NGS – Pipeline Maison






```
#########################
#       Kiwi v1.0       #
#  Alignment Optimizer  #
#########################


   Start time: 18/12/15 09:38:50
   Input BAM :  /home/data/Auto_user_Proton-176-RUN022_ONCO_272_429/IonXpress_048/IonXpress_048.bam
   Output BAM: /home/results/Auto_user_Proton-176-RUN022_ONCO_272_429/Kiwi/IonXpress_048_Kiwi.bam
   GATK RealignerTargetCreator [OK]
   GATK IndelRealigner         [OK]
   GATK BaseRecalibrator       [OK]
   GATK PrintReads             [OK]
   Read/Update NkGene Database [OK]
   Pseudogene analysis
         <Create/Read BED> (31 targets with pseudogene)
         <Mask Reference (0)>
         <Reduce BAM (0)>
         <BAM to FastQ (0)>
         <Bowtie2 mapping (0)>
         <Sort & Index BAM (0)>
   Read WellPlateDataSheet     [OK]
   Create Kiwi BAMs            [OK]
   Create Kiwi Pseudo BAMs     [OK]
   Index Kiwi BAM              [OK]
   Validate BAM file           [OK]
   Clean temporary folder      [OK]
   End time  : 18/12/15 11:05:23
   SUCCESS
```

Realignment / Recalibration

Pseudogene Analysis (NkPseudo)

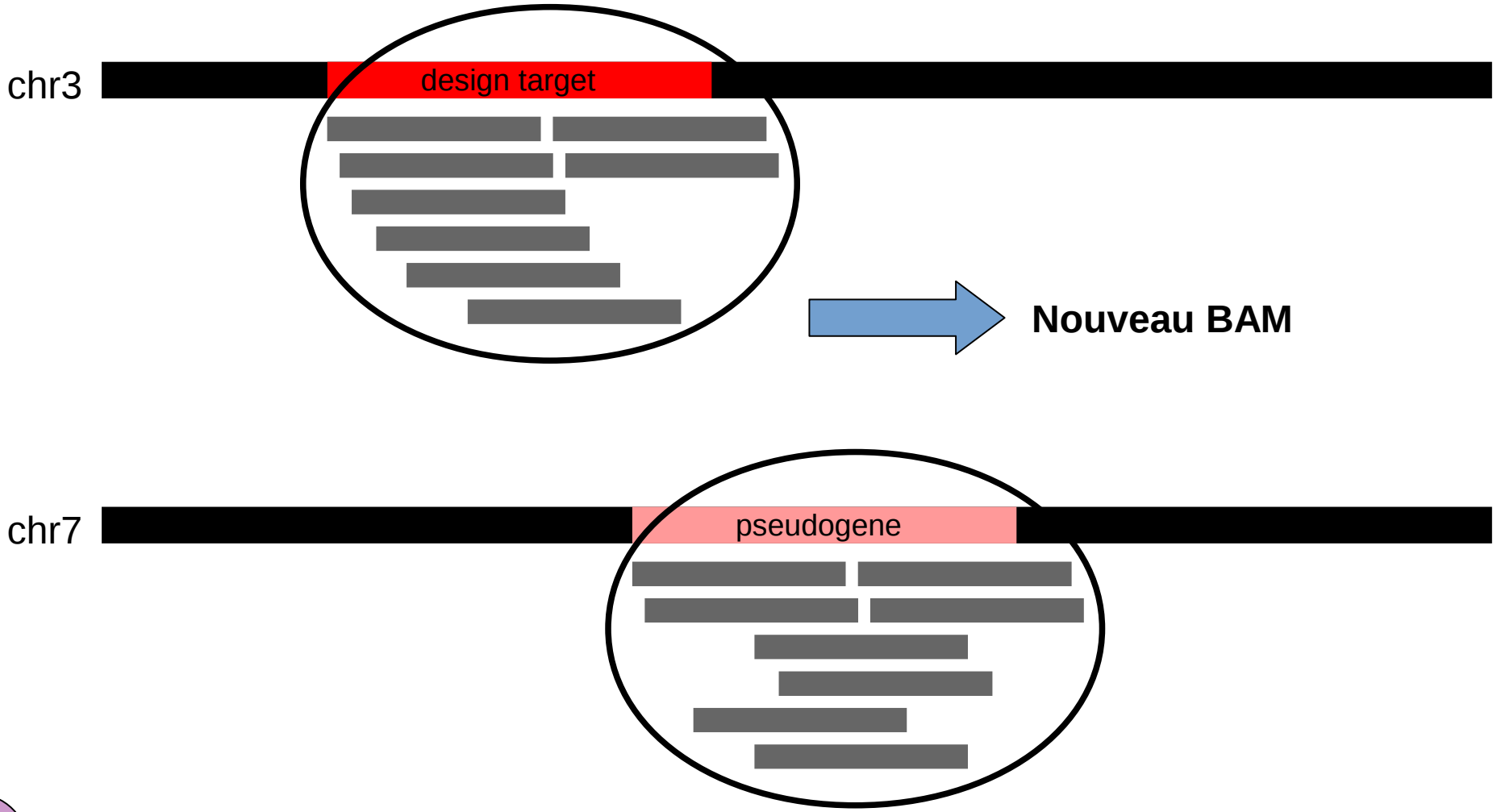
Small primer reads filtering / q>20

NkPseudo - Pseudogene analysis

1. Determiner les régions du design avec des pseudogènes (blastn sur hg19)
   Seuil d'identité de 95 %

NkPseudo - Pseudogene analysis

2 Récupération des reads mappés sur « target » et « pseudogene »



NkPseudo

3 Désalignement des reads => Conversion du BAM en FastQ

>read1

>read2

>read3

>read4

>read5

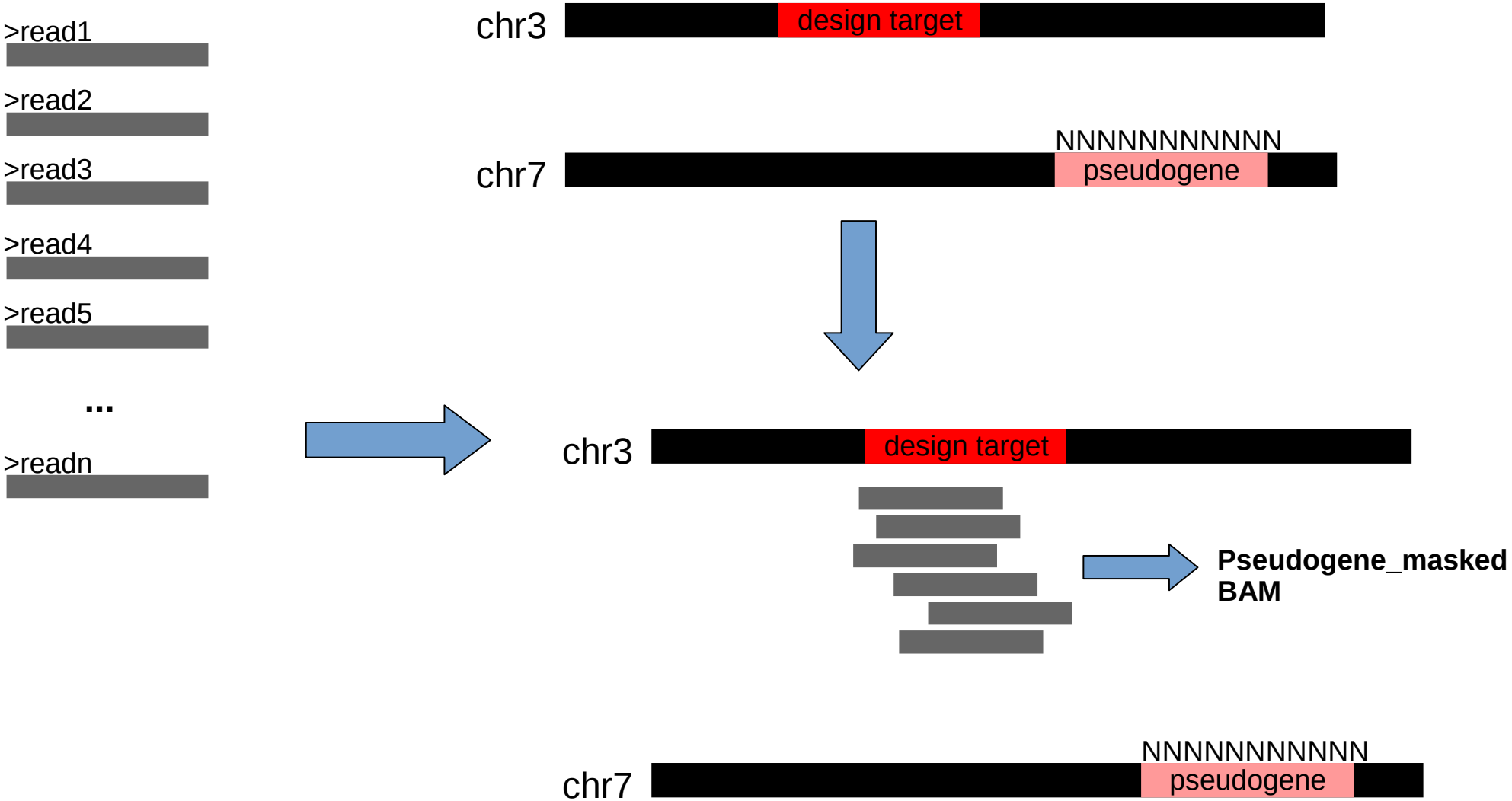
...

>readn

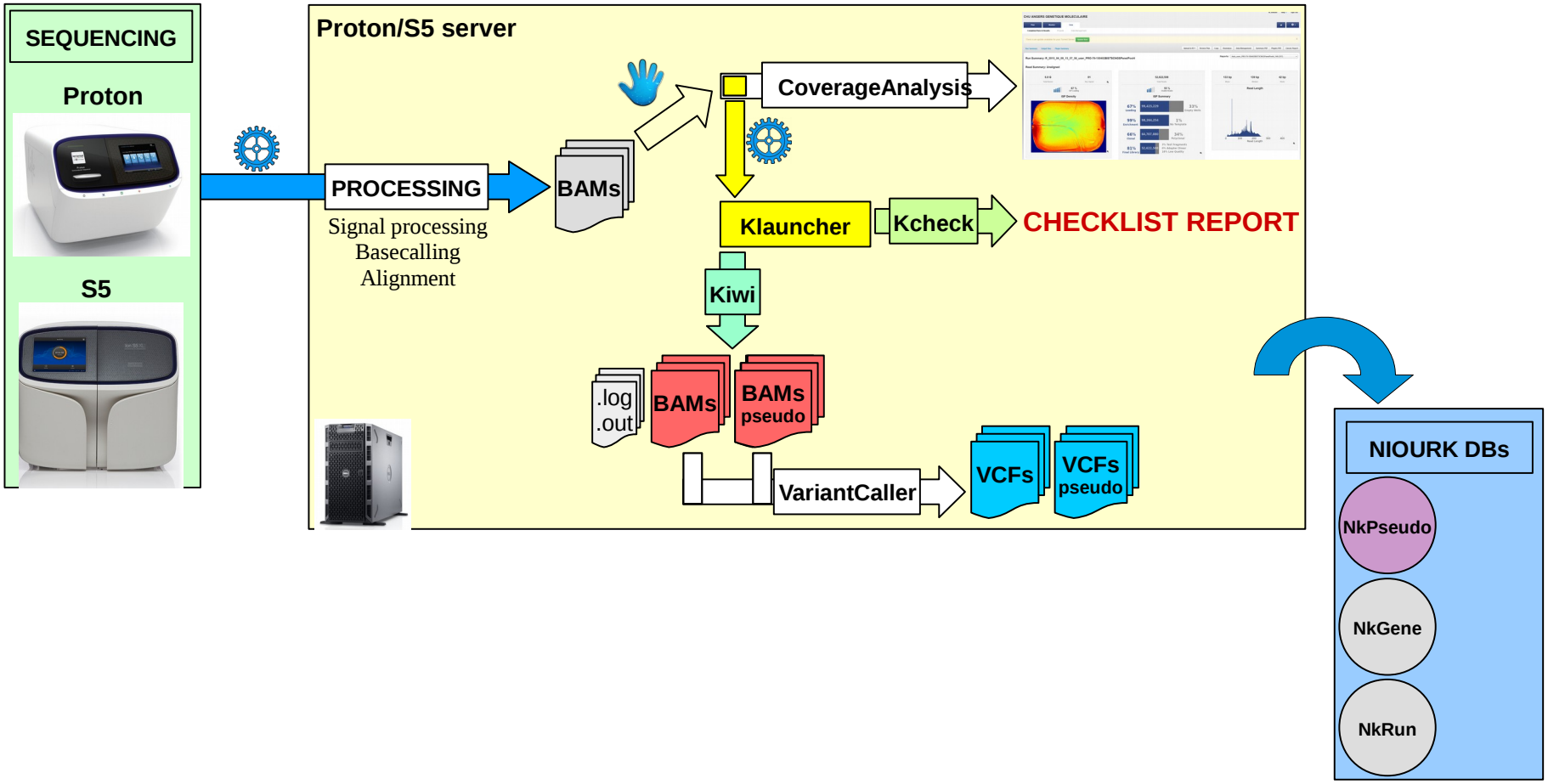
4 Masquage des pseudogènes sur la référence (hg19)

NkPseudo - Pseudogene analysis

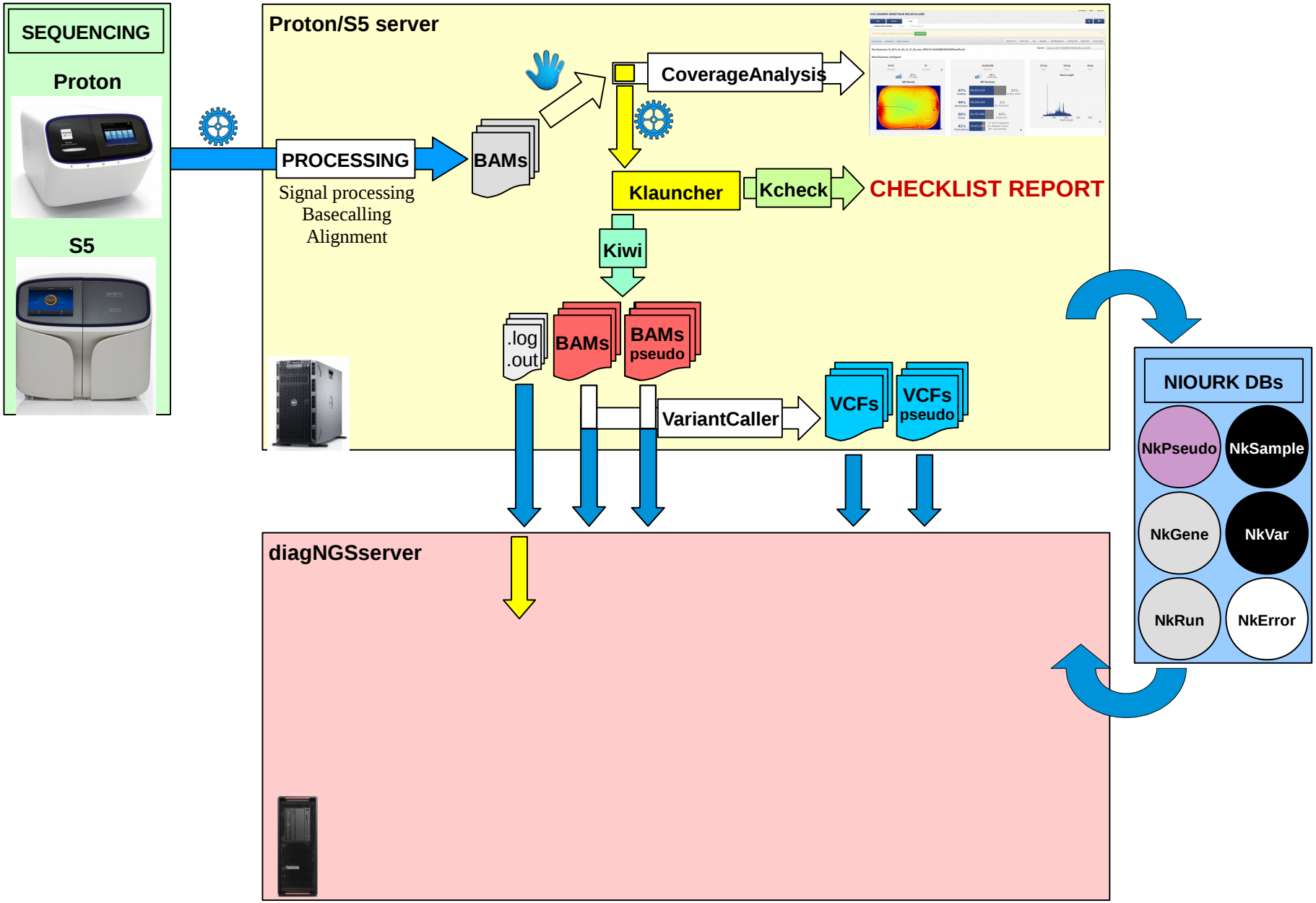
5 Réalignement sur la référence masquée avec bowtie2



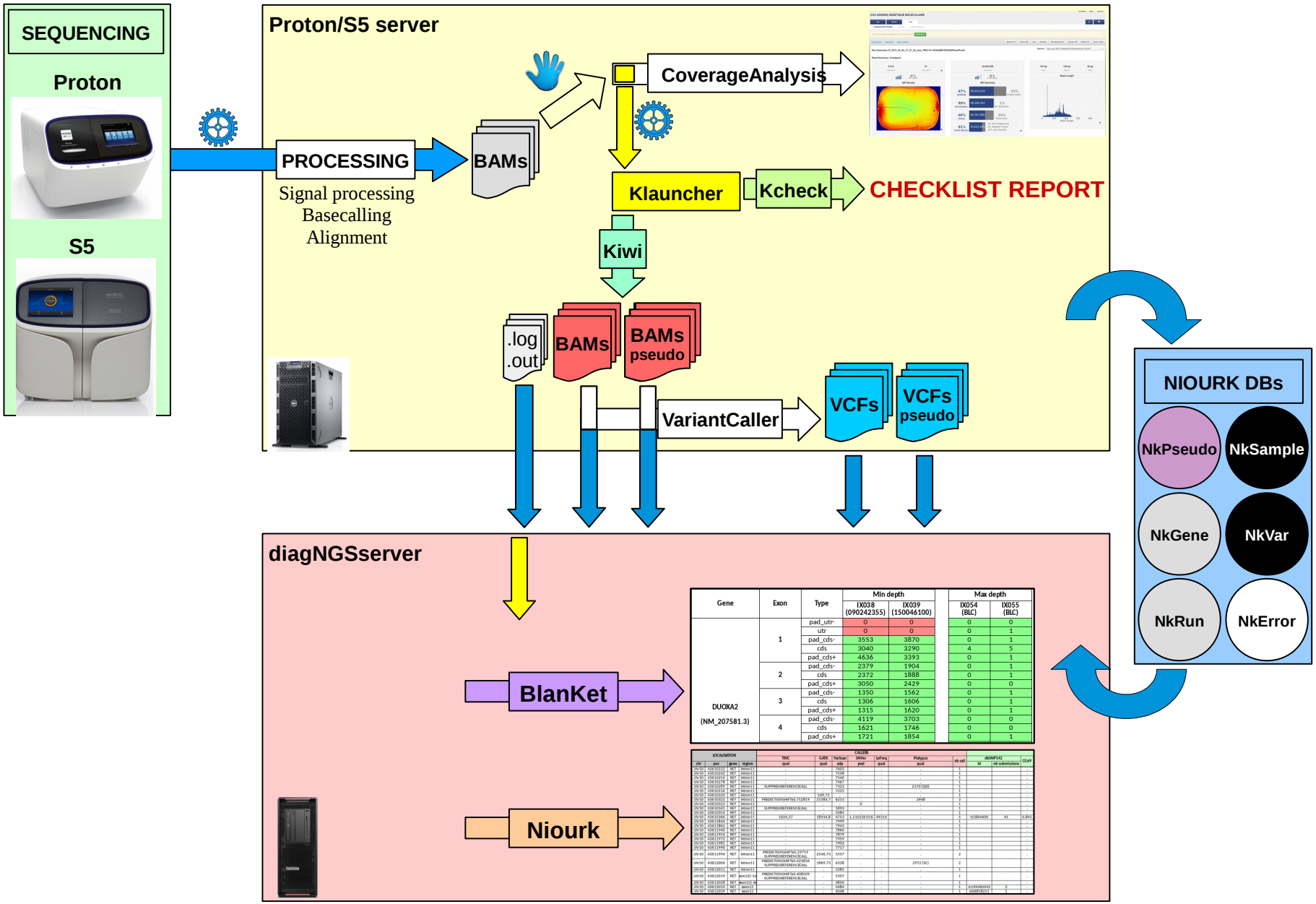
Diagnostique NGS – Pipeline Maison



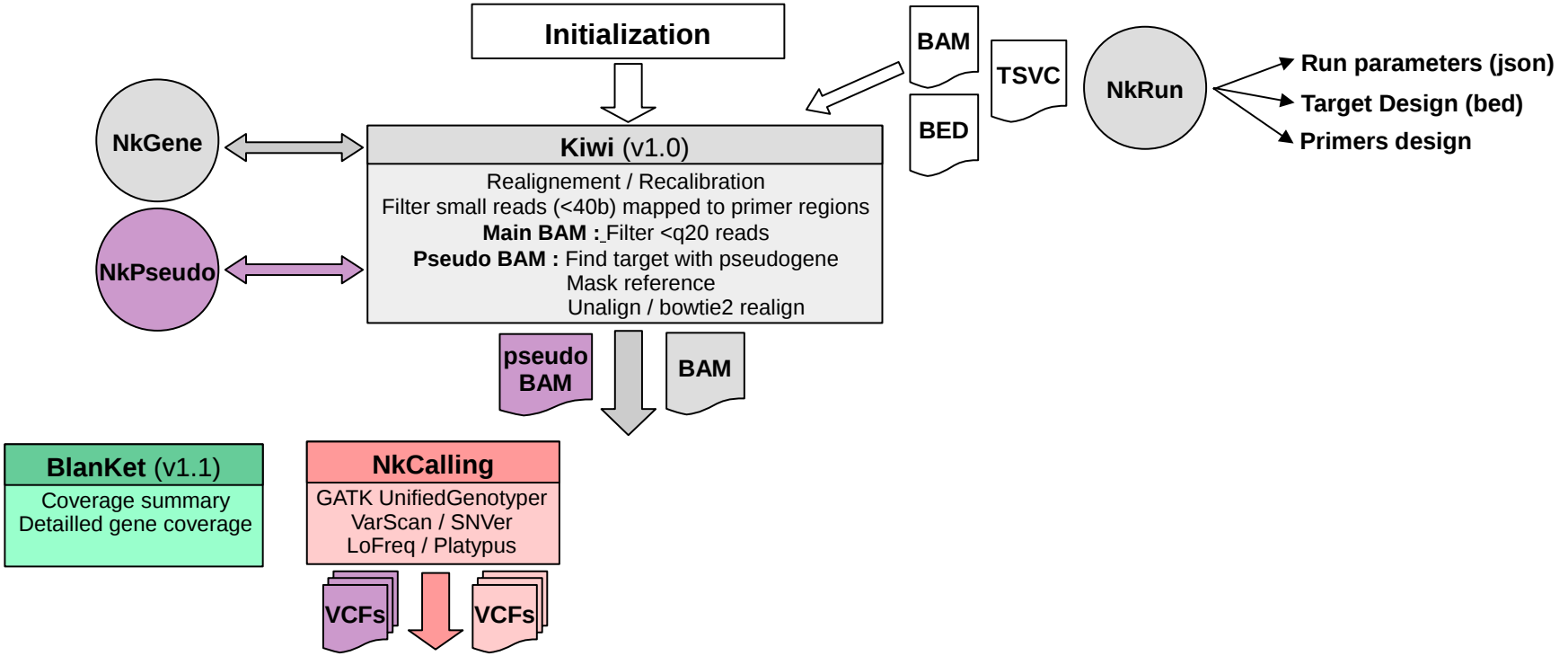
Diagnostique NGS – Pipeline Maison



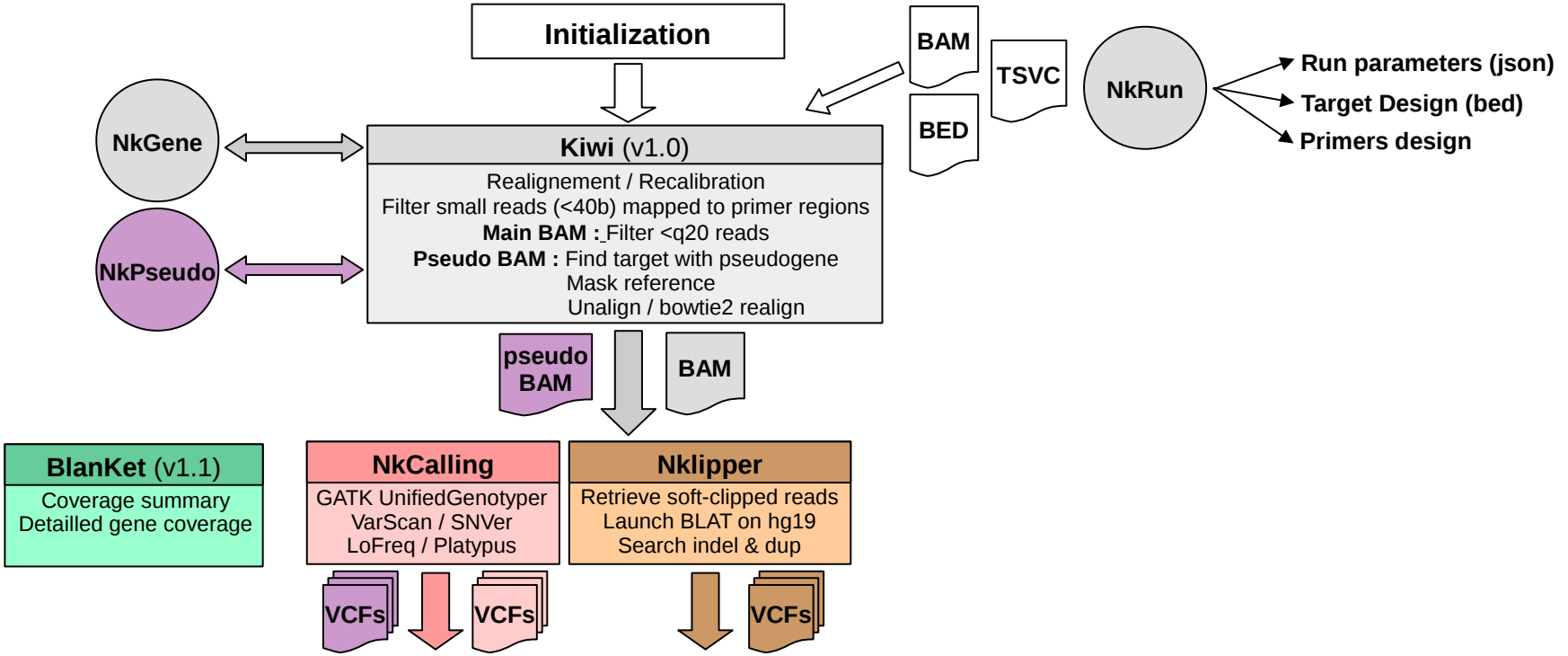
Diagnostique NGS – Pipeline Maison



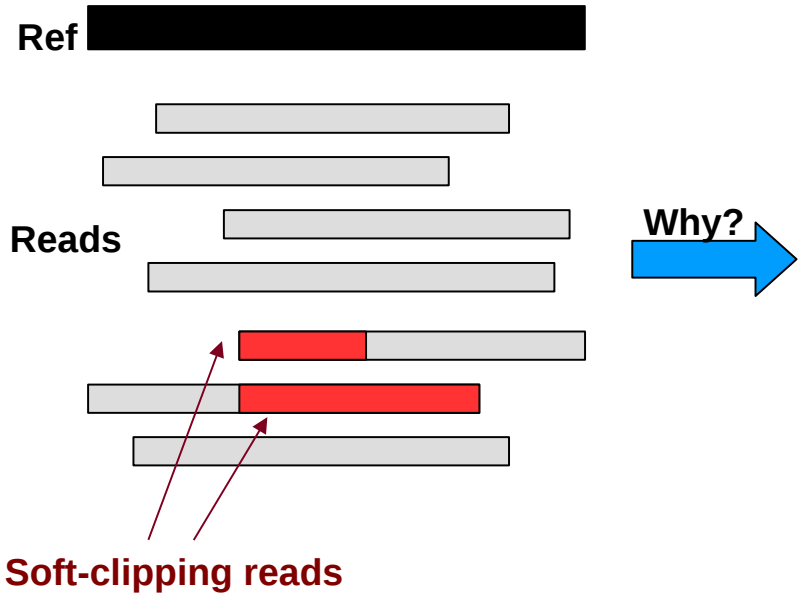
Niourk v1.7
Variant Calling Pipeline



Niourk v1.7
Variant Calling Pipeline

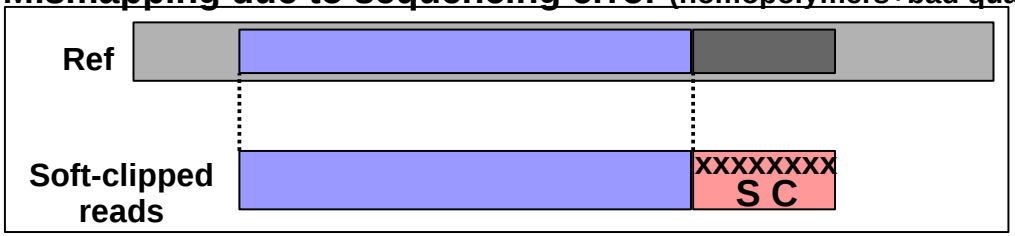


Nklipper – Soft-clipping analysis

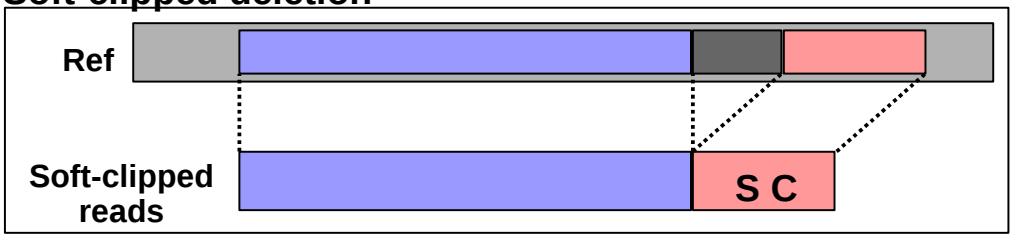


Why?

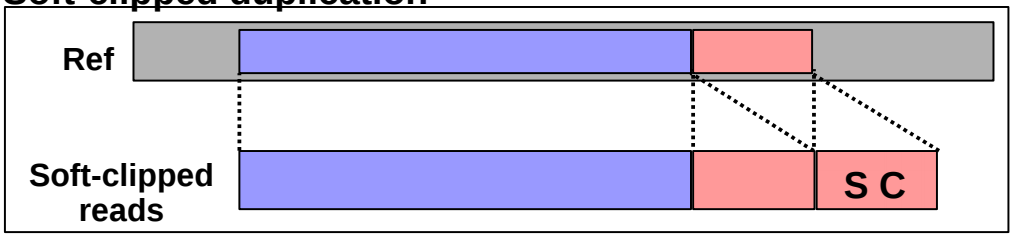
Mismapping due to sequencing error (homopolymers+bad quality)



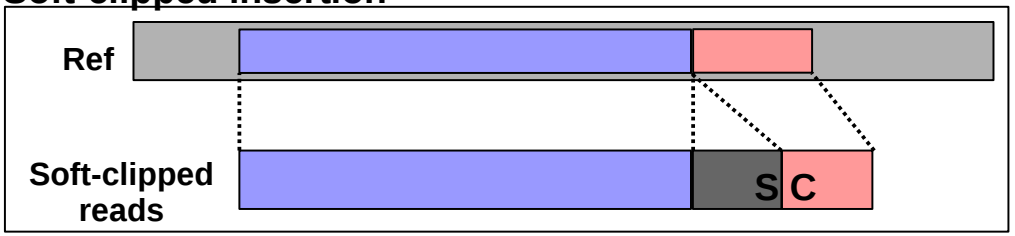
Soft-clipped deletion



Soft-clipped duplication



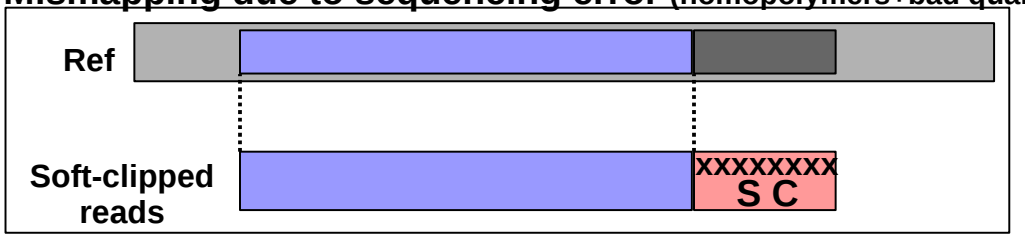
Soft-clipped insertion



Nklipper – Soft-clipping analysis

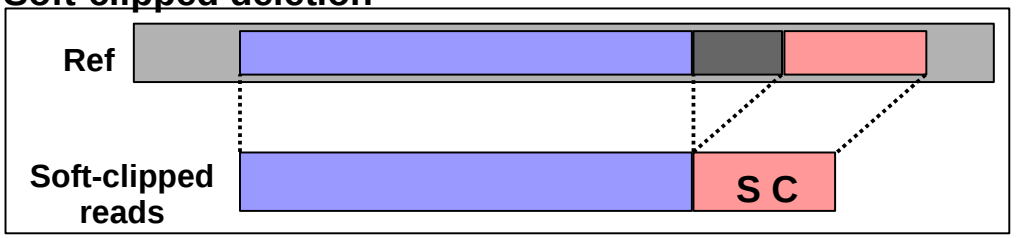
Bad read end quality or reference/read homopolymers

Mismapping due to sequencing error (homopolymers+bad quality)



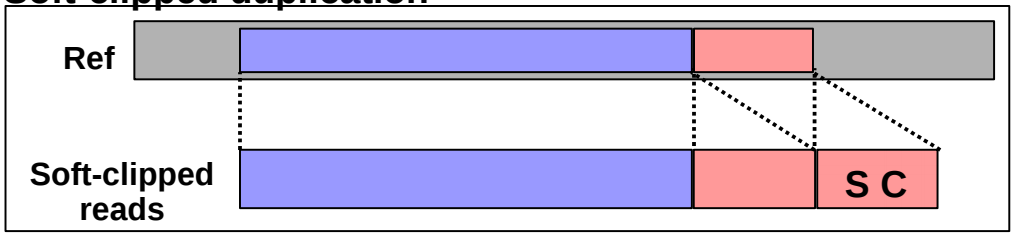
Soft-clipped sequence could be upstream mapped on reference

Soft-clipped deletion



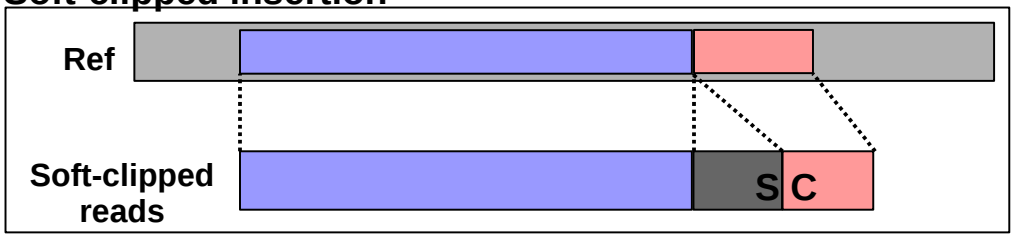
Soft-clipped sequence could be normal mapped on reference

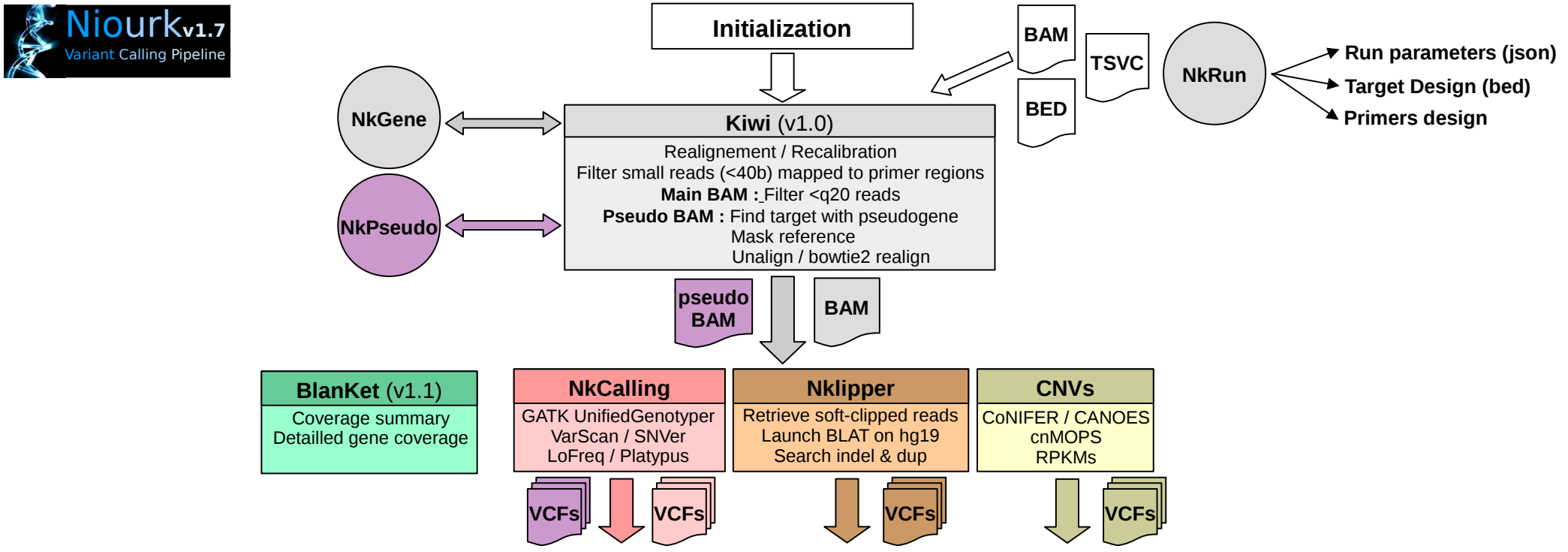
Soft-clipped duplication

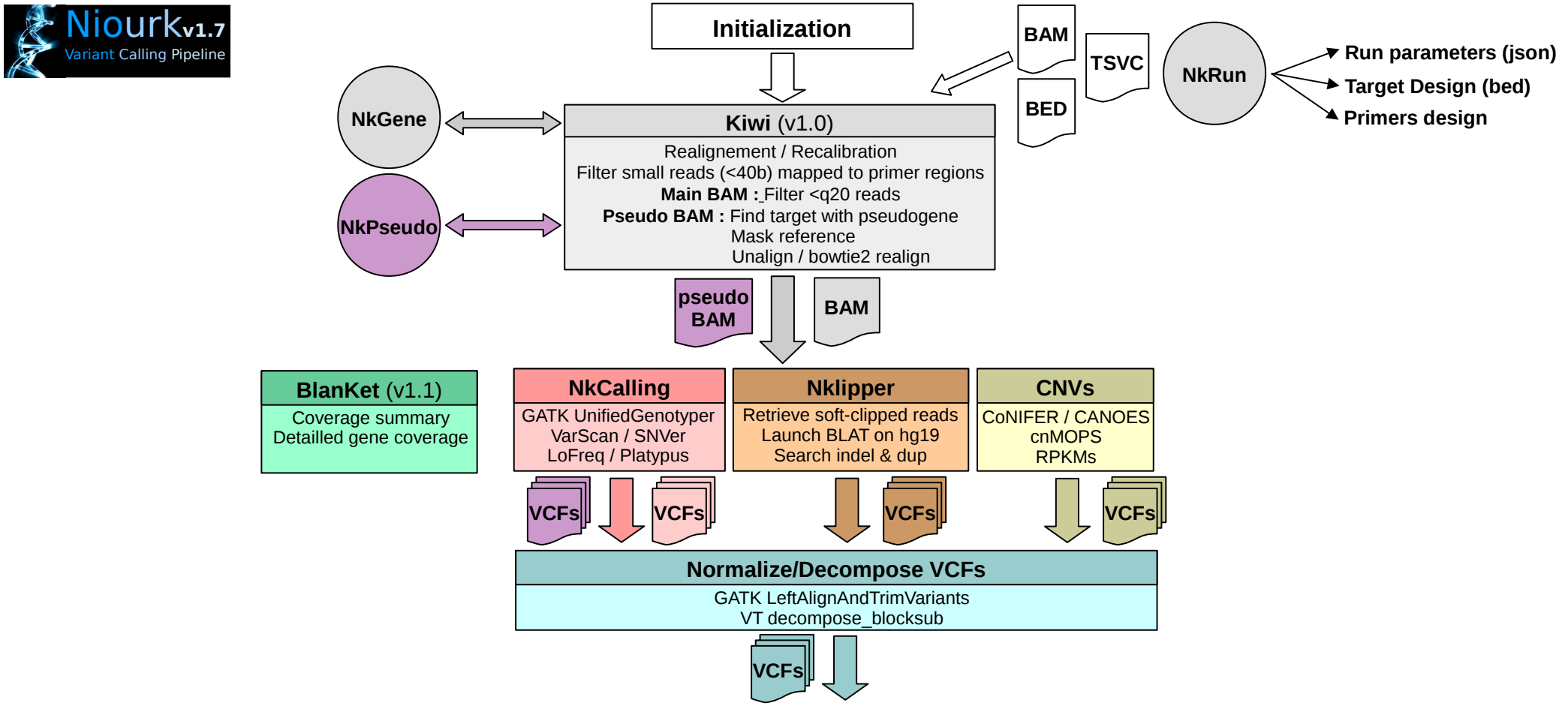


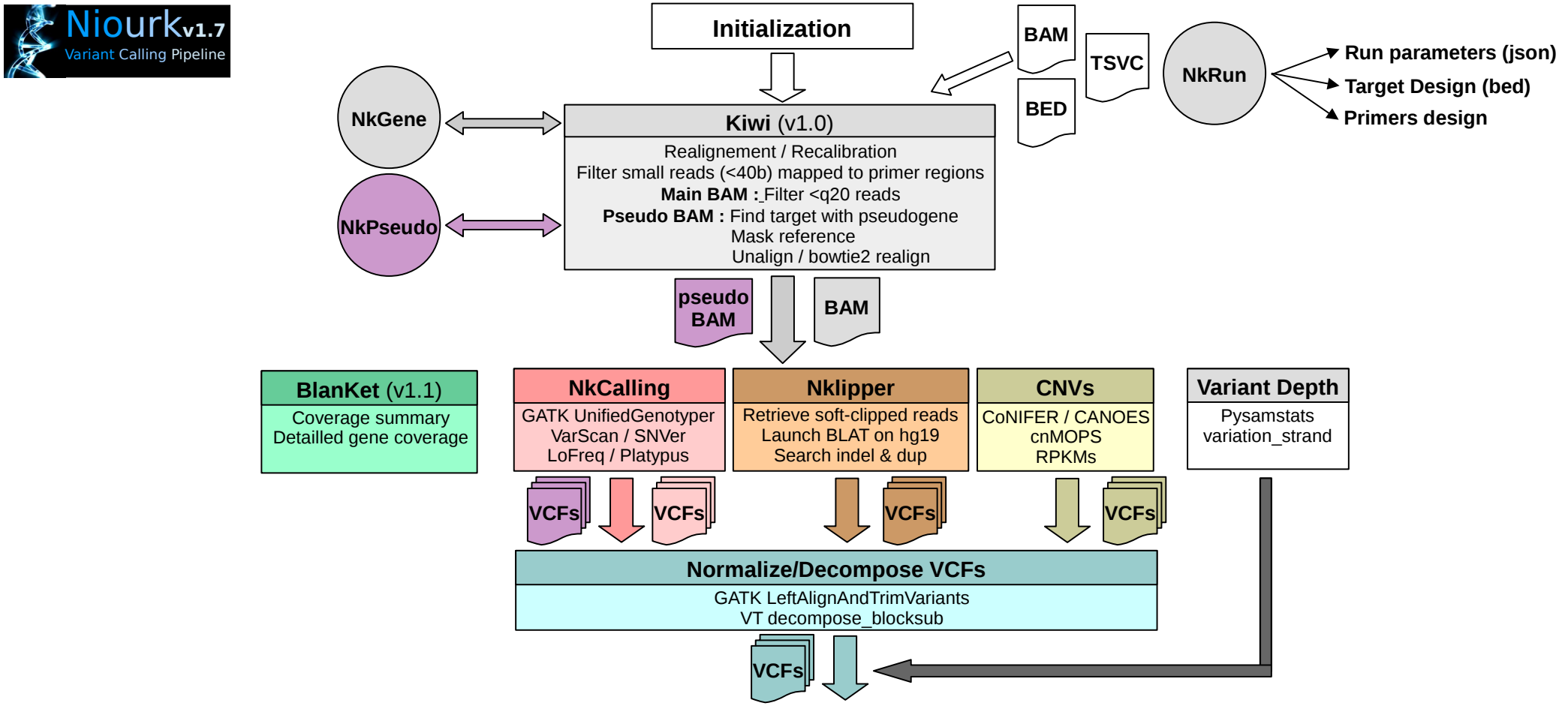
Soft-clipped sequence could be downstream mapped

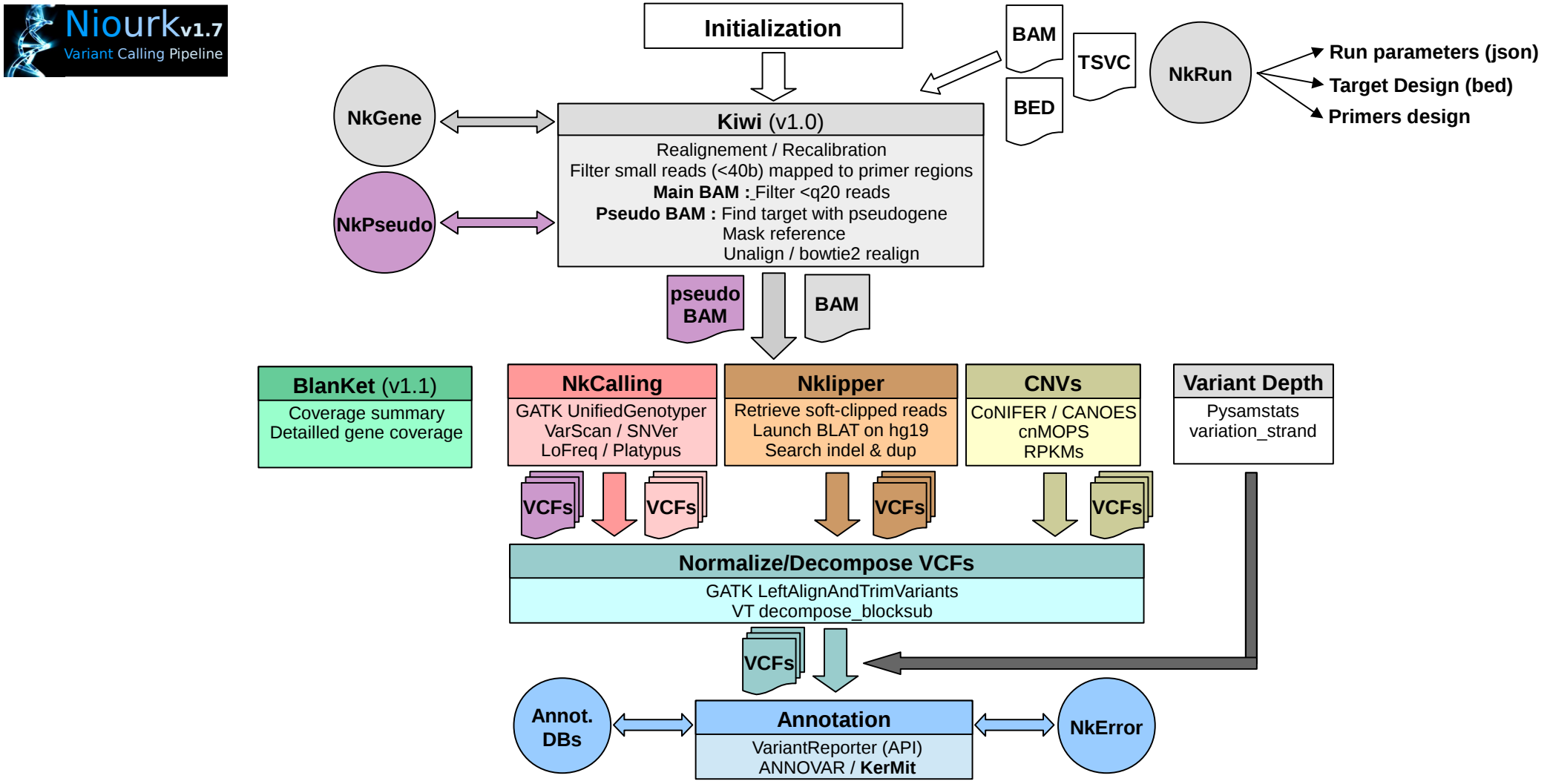
Soft-clipped insertion



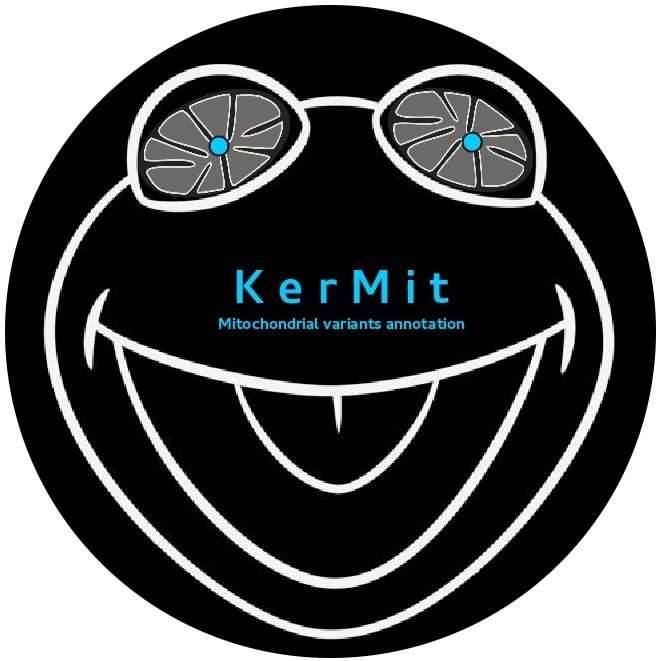

Niourk v1.7
Variant Calling Pipeline
Initialization
BAM
TSVC
BED
NkRun
Run parameters (json)
Target Design (bed)
Primers design
NkGene
NkPseudo
Kiwi (v1.0)
Realignement / Recalibration
Filter small reads (<40b) mapped to primer regions
Main BAM : Filter <q20 reads
Pseudo BAM : Find target with pseudogene
Mask reference
Unalign / bowtie2 realign
pseudo BAM
BAM
BlanKet (v1.1)
Coverage summary
Detailled gene coverage
NkCalling
GATK UnifiedGenotyper
VarScan / SNVer
LoFreq / Platypus
Nklipper
Retrieve soft-clipped reads
Launch BLAT on hg19
Search indel & dup
CNVs
CoNIFER / CANOES
cnMOPS
RPKMs
VCFs
VCFs
VCFs
VCFs


Niourk v1.7
Variant Calling Pipeline
Initialization
BAM
TSVC
BED
NkRun
Run parameters (json)
Target Design (bed)
Primers design
NkGene
NkPseudo
Kiwi (v1.0)
Realignement / Recalibration
Filter small reads (<40b) mapped to primer regions
Main BAM : Filter <q20 reads
Pseudo BAM : Find target with pseudogene
Mask reference
Unalign / bowtie2 realign
pseudo BAM
BAM
BlanKet (v1.1)
Coverage summary
Detailled gene coverage
NkCalling
GATK UnifiedGenotyper
VarScan / SNVer
LoFreq / Platypus
Nklipper
Retrieve soft-clipped reads
Launch BLAT on hg19
Search indel & dup
CNVs
CoNIFER / CANOES
cnMOPS
RPKMs
VCFs
VCFs
VCFs
VCFs
Normalize/Decompose VCFs
GATK LeftAlignAndTrimVariants
VT decompose_blocksub
VCFs


Niourk v1.7
Variant Calling Pipeline
Initialization
BAM
TSVC
BED
NkRun
Run parameters (json)
Target Design (bed)
Primers design
NkGene
NkPseudo
Kiwi (v1.0)
Realignement / Recalibration
Filter small reads (<40b) mapped to primer regions
Main BAM : Filter <q20 reads
Pseudo BAM : Find target with pseudogene
Mask reference
Unalign / bowtie2 realign
pseudo BAM
BAM
BlanKet (v1.1)
Coverage summary
Detailled gene coverage
NkCalling
GATK UnifiedGenotyper
VarScan / SNVer
LoFreq / Platypus
Nklipper
Retrieve soft-clipped reads
Launch BLAT on hg19
Search indel & dup
CNVs
CoNIFER / CANOES
cnMOPS
RPKMs
Variant Depth
Pysamstats
variation_strand
VCFs
VCFs
VCFs
VCFs
Normalize/Decompose VCFs
GATK LeftAlignAndTrimVariants
VT decompose_blocksub
VCFs

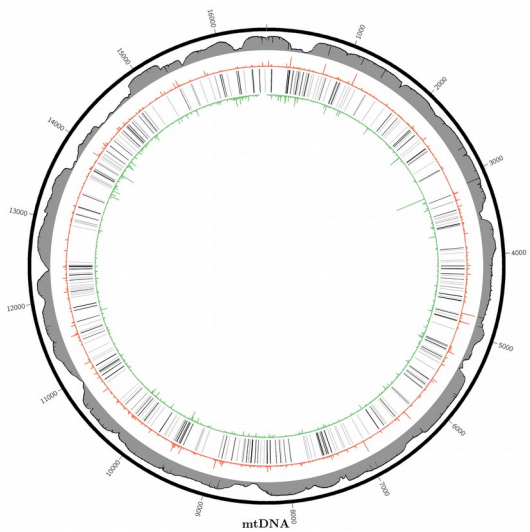

Niourk v1.7
Variant Calling Pipeline
Initialization
BAM
TSVC
BED
NkRun
Run parameters (json)
Target Design (bed)
Primers design
NkGene
NkPseudo
Kiwi (v1.0)
Realignement / Recalibration
Filter small reads (<40b) mapped to primer regions
Main BAM : Filter <q20 reads
Pseudo BAM : Find target with pseudogene
Mask reference
Unalign / bowtie2 realign
pseudo BAM
BAM
BlanKet (v1.1)
Coverage summary
Detailled gene coverage
NkCalling
GATK UnifiedGenotyper
VarScan / SNVer
LoFreq / Platypus
Nklipper
Retrieve soft-clipped reads
Launch BLAT on hg19
Search indel & dup
CNVs
CoNIFER / CANOES
cnMOPS
RPKMs
Variant Depth
Pysamstats
variation_strand
VCFs
VCFs
VCFs
VCFs
Normalize/Decompose VCFs
GATK LeftAlignAndTrimVariants
VT decompose_blocksub
VCFs
Annot. DBs
Annotation
VariantReporter (API)
ANNOVAR / KerMit
NkError

KerMiT – Mitochondrial variant annotation

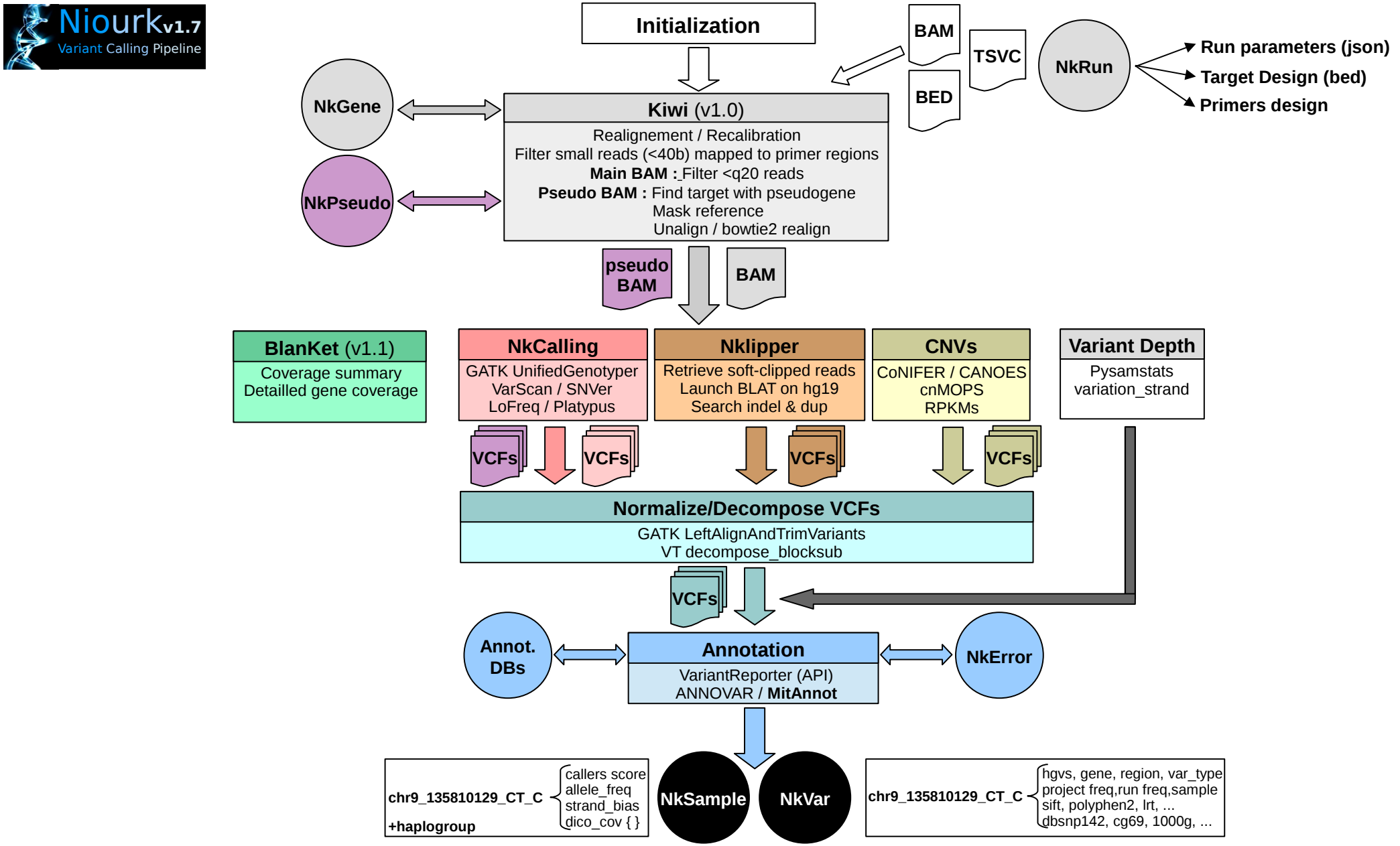


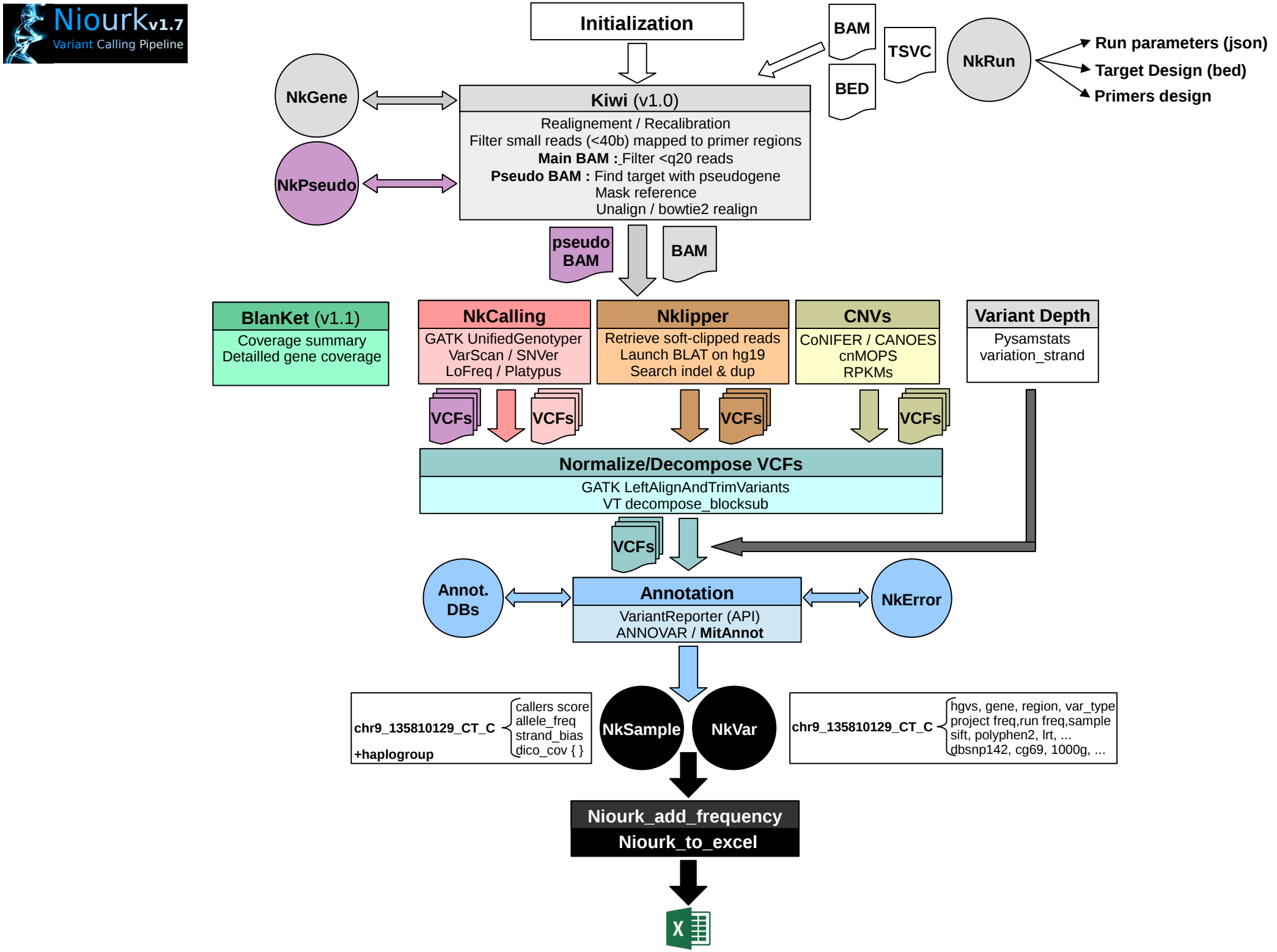
- Définition de l'haplogroupe avec phymer (a. p. du BAM)
- Elimination des variants définissant cet haplogroupe
- Utilisation de l'API Mitomaster
- Utilisation de MitImpact 2.4
- DB KerMit (données + métadonnées)
- Visualisation Circos

...



En cours ...


Niourk v1.7
Variant Calling Pipeline
Initialization
BAM
TSVC
BED
NkRun
Run parameters (json)
Target Design (bed)
Primers design
NkGene
NkPseudo
Kiwi (v1.0)
Realignement / Recalibration
Filter small reads (<40b) mapped to primer regions
Main BAM : Filter <q20 reads
Pseudo BAM : Find target with pseudogene
Mask reference
Unalign / bowtie2 realign
pseudo BAM
BAM
BlanKet (v1.1)
Coverage summary
Detailled gene coverage
NkCalling
GATK UnifiedGenotyper
VarScan / SNVer
LoFreq / Platypus
Nklipper
Retrieve soft-clipped reads
Launch BLAT on hg19
Search indel & dup
CNVs
CoNIFER / CANOES
cnMOPS
RPKMs
Variant Depth
Pysamstats
variation_strand
VCFs
VCFs
VCFs
VCFs
Normalize/Decompose VCFs
GATK LeftAlignAndTrimVariants
VT decompose_blocksub
VCFs
Annot. DBs
Annotation
VariantReporter (API)
ANNOVAR / MitAnnot
NkError
chr9_135810129_CT_C
+haplogroup
callers score
allele_freq
strand_bias
dico_cov { }
NkSample
NkVar
chr9_135810129_CT_C
hgvs, gene, region, var_type
project freq,run freq,sample
sift, polyphen2, lrt, ...
dbsnp142, cg69, 1000g, ...


Niourk v1.7
Variant Calling Pipeline
Initialization
BAM
TSVC
BED
NkRun
Run parameters (json)
Target Design (bed)
Primers design
NkGene
NkPseudo
Kiwi (v1.0)
Realignement / Recalibration
Filter small reads (<40b) mapped to primer regions
Main BAM : Filter <q20 reads
Pseudo BAM : Find target with pseudogene
Mask reference
Unalign / bowtie2 realign
pseudo BAM
BAM
BlanKet (v1.1)
Coverage summary
Detailled gene coverage
NkCalling
GATK UnifiedGenotyper
VarScan / SNVer
LoFreq / Platypus
Nklipper
Retrieve soft-clipped reads
Launch BLAT on hg19
Search indel & dup
CNVs
CoNIFER / CANOES
cnMOPS
RPKMs
Variant Depth
Pysamstats
variation_strand
VCFs
VCFs
VCFs
VCFs
Normalize/Decompose VCFs
GATK LeftAlignAndTrimVariants
VT decompose_blocksub
VCFs
Annot. DBs
Annotation
VariantReporter (API)
ANNOVAR / MitAnnot
NkError
chr9_135810129_CT_C
+haplogroup
callers score
allele_freq
strand_bias
dico_cov { }
NkSample
NkVar
chr9_135810129_CT_C
hgvs, gene, region, var_type
project freq,run freq,sample
sift, polyphen2, lrt, ...
dbsnp142, cg69, 1000g, ...
Niourk_add_frequency
Niourk_to_excel
X

Niourk v1.7
Variant Calling Pipeline

```
################################
#          NIOURK v1.7         #
#   Variant Calling Pipeline   #
################################
```

```
Initialization
  start time:21/12/15 15:44  run:Auto_user_Proton-176-RUN022_ONCO_272_429
  bam:IonXpress_012    target:Pheo_CustomPanel_IAD59495_182_Designed.bed
  min_cov:20   min_base_q:15   min_read_q:20   min_alt_freq:0.02   strand_bias:0.95   pval:0.05
Read/Update NkGene Database [OK]
```

Initialization

```
GATK UnifiedGenotyper       [OK]
VarScan mpileup2cns         [OK]
SNVerIndividual             [OK]
LoFreq Call                 [OK]
Platypus callVariants       [OK]
SCP TSVC to diagNGSserver   [OK]
Pseudogene calling          [OK]
```

Calling

```
Normalize/Decompose VCFs    [OK]
VariantReporter annotation  [OK]
Add variant calling results [OK]
```

Variants analysis

```
Convert/Write ANNOVAR input [OK]
ANNOVAR Variant Annotation  [OK]
 ANNOVAR:table_annovar.pl(2015-12-14)   refGene:(2015-12-11)   COSMIC:v70 (2014-09-11)
  dbscSNV:v1.1(2015-12-18)   1000G:ALL(2015-08-24)   ExAC:65000 exome v0.3(2015-11-29)
  NHLBI-ESP:6500siv2(2014-12-22) CG:69(2012-02-22) NCI60:(2013-07-24) dbnsfp31a_interpro:(2015-12-19)
  dbSNP&ClinVar:via VariantReporterAPI   kaviar:(2015-12-03)   hrcr1:(2015-12-03)
  SIFT,PolyPhen2,LRT,MutationTaster,MutationAssessor,...,SiPhy:v3.0a(2015-10-15)
```

Annotation (ANNOVAR)

```
Variants coverage analysis  [OK]
```

Pysamstats

```
Fill Niourk Database        [OK]
Write Niourk JSON           [OK]
```

Write in DBs

```
Clean temp files            [OK]
End time  : 21/12/15 15:44:56
SUCCESS
```

Tableau final

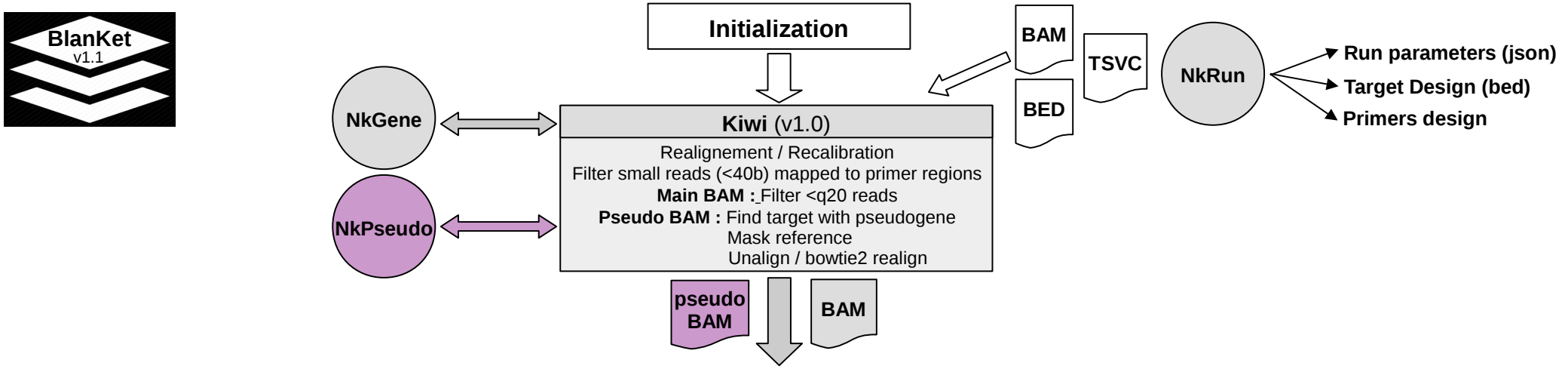
LOCALISATION					VARIANT							
chr	pos	gene	region	pseudo	ref	alt	hgvs_c	hgvs_p	link	type	user annot	comments

FEATURES						DEPTH																				
Allele Freq	Relative Sb	frequency		frequency > 4call		Total			A			T			G			C			ins			del		
		project	run	project	run	all	fwd	rev	all	fwd	rev	all	fwd	rev	all	fwd	rev	all	fwd	rev	all	fwd	rev	all	fwd	rev

CALLERS						
TSVC	GATK	VarScan	SNVer	LoFreq	Platypus	nb call
qual	qual	adp	pval	qual	qual	

DATABASES																						
dbSNP144		COSMIC70		CLINVAR				Max DBs freq	CG69	NCI60	ESP	EXAC03								Kaviar	HRC
id	occurence	id	occurence	accession	significance	evidence	phenotype					all	afr	amr	eas	fin	nfe	oth	sas		

PRIORIZATION TOOLS													
SIFT	Polyphen2		LRT	MutationTaster	MutationAssessor	FATHMM		PROVEAN	MetaSVM	MetaLR	Interpro	dbscSNV	
	HDIV	HVAR				pred	MKL coding pred					ADA score	RF score

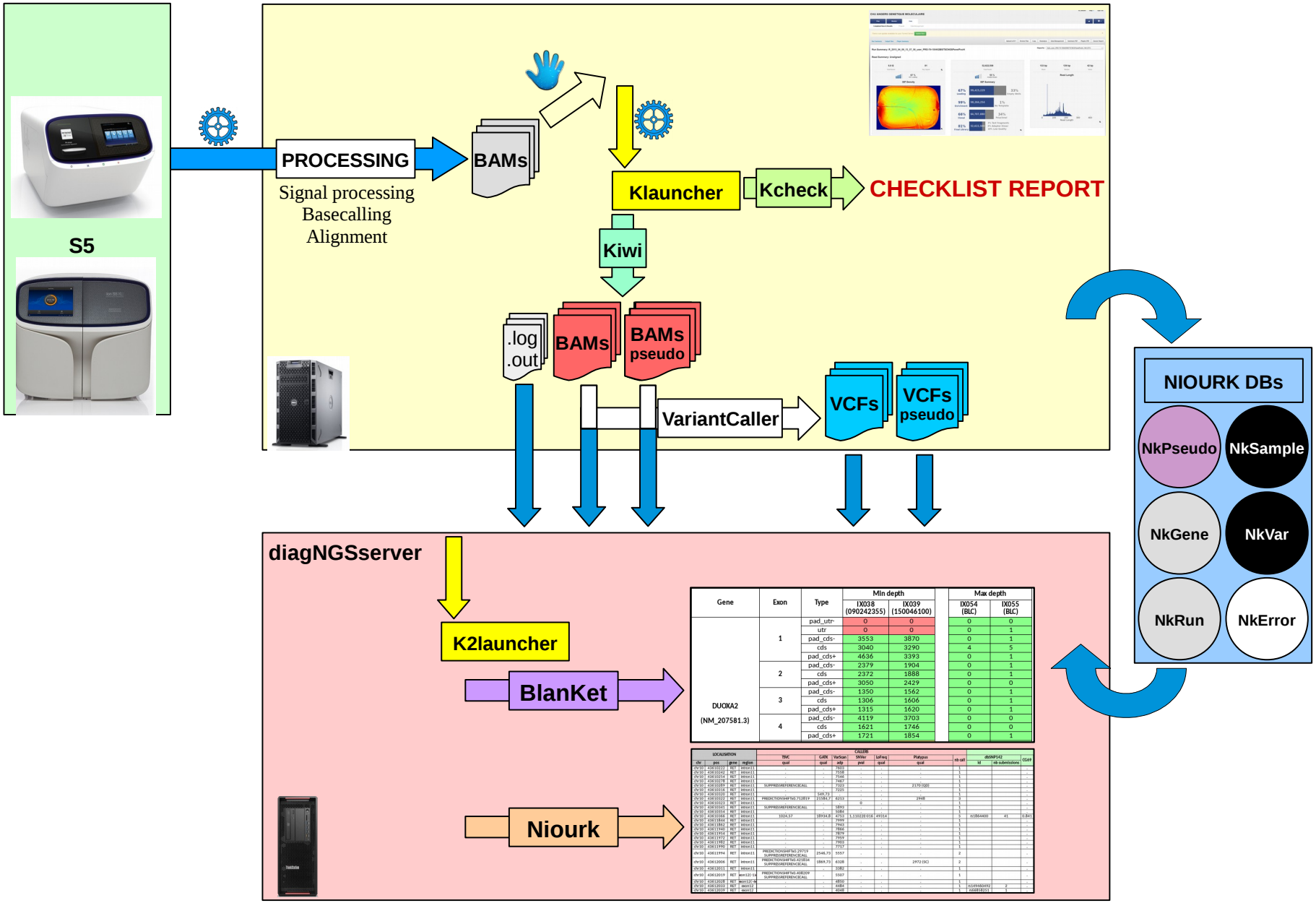



```
##########################################
# BlanKet v1.1 (NGS Coverage analysis) #
##########################################

  Initialization                [OK]
  Read Target Regions (bed)     [OK]  (DUOXA1,IYD,(...),TSHB,SERPINA7)
  Read gene_to_transcript file  [OK]
  Retrieve genes features       [OK]          Genomic organization formatting
  Convert genomic positions     [OK]
  Write multi-genes bed file    [OK]
  Launch coverageBed                  (bedtools v2.25.0)
        IonXpress_038           [OK]
        IonXpress_038 (pseudo)  [OK]
        IonXpress_039           [OK]
        IonXpress_039 (pseudo)  [OK]                      Coverage computing
        (...)
        IonXpress_054           [OK]
        IonXpress_054 (pseudo)  [OK]
        IonXpress_055           [OK]
        IonXpress_055 (pseudo)  [OK]
  Create summary XLS table      [OK]
  Create genes XLS table        [OK]          Excel table creation
  Clean temporary folder        [OK]
```

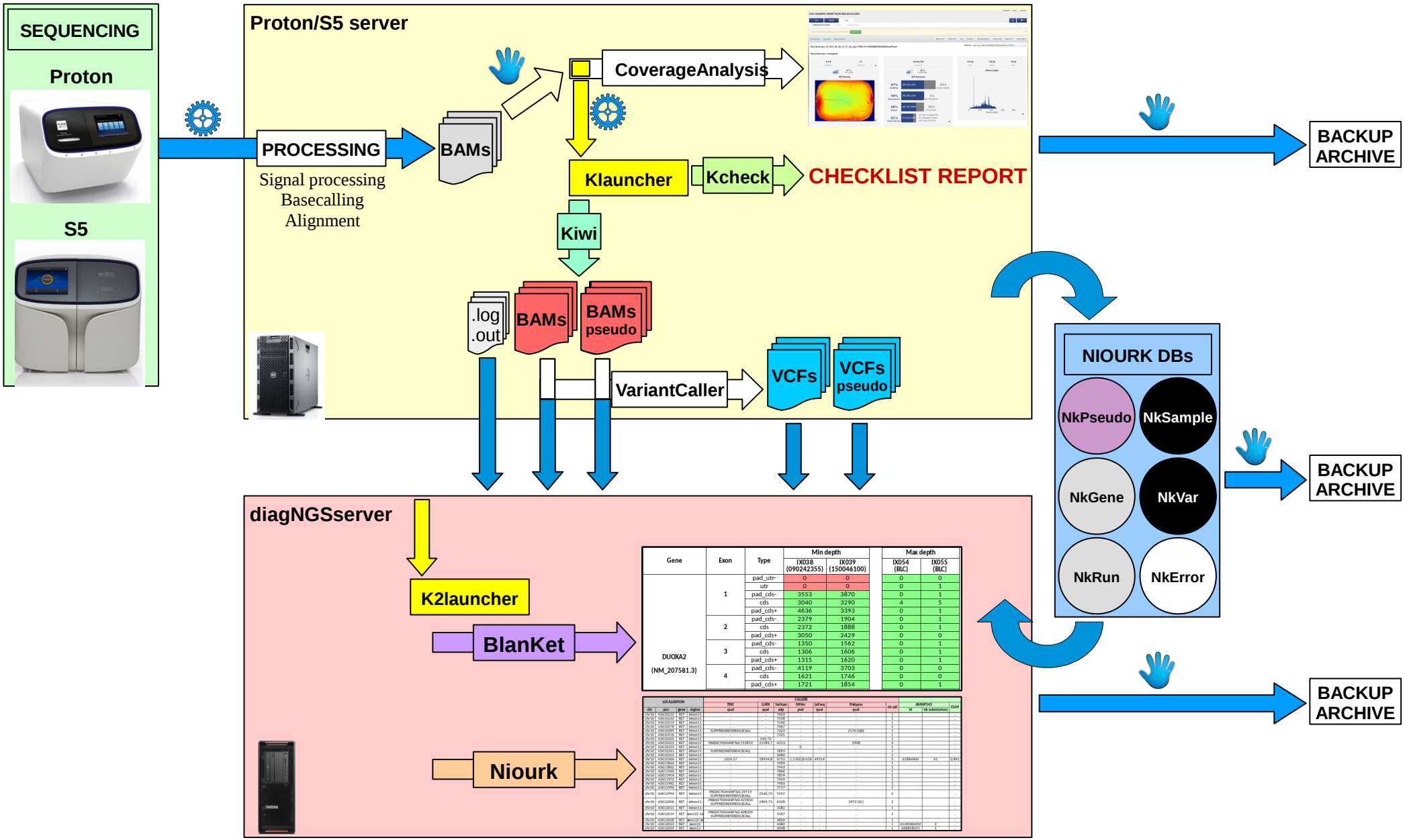
Summary table

Gene	Exon	Type	Min depth		Max depth	
			IX038 (090242355)	IX039 (150046100)	IX054 (BLC)	IX055 (BLC)
DUOXA2 (NM_207581.3)	1	pad_utr-	0	0	0	0
		utr	0	0	0	1
		pad_cds-	3553	3870	0	1
		cds	3040	3290	4	5
		pad_cds+	4636	3393	0	1
	2	pad_cds-	2379	1904	0	1
		cds	2372	1888	0	1
		pad_cds+	3050	2429	0	0
	3	pad_cds-	1350	1562	0	1
		cds	1306	1606	0	1
		pad_cds+	1315	1620	0	1
	4	pad_cds-	4119	3703	0	0
		cds	1621	1746	0	0
		pad_cds+	1721	1854	0	1
	5	pad_cds-	506	419	0	1
		cds	479	400	0	0
		pad_cds+	2842	2063	0	0
	6	utr	55	102	0	0
		pad_utr+	0	0	0	0
		pad_cds-	5115	4488	0	0
		cds	1441	2211	0	0
		pad_cds+	1468	2266	0	1
DUOX1 (NM_017434.4)	1	pad_utr-	0	0	0	0
		utr	0	0	0	0
		pad_utr+	0	0	0	0
	2	pad_utr-	0	0	0	0
		utr	0	0	0	0
		pad_utr+	0	0	0	0
	3	pad_utr-	858	1057	193	493
		utr	840	1045	193	494
		pad_cds-	831	1037	191	483
		cds	803	1027	189	479
		pad_cds+	786	1006	173	463
	4	pad_cds-	4614	4168	0	1
		cds	3549	3038	0	1
		pad_cds+	4213	2962	0	0

Diagnostique NGS – Pipeline Maison



Diagnostique NGS – Pipeline Maison



Ou récupérer les résultats : http://172.29.32.118/run_results/

Index of /run_results

Name	Last modified	Size	Description
Parent Directory		-	
Auto_user_PRO-101-15..>	2016-01-04 12:09	-	
Auto_user_PRO-102-15..>	2016-01-04 04:58	-	
Auto_user_PRO-105-15..>	2016-01-04 01:02	-	
Auto_user_PRO-106-15..>	2016-01-03 21:50	-	
Auto_user_PRO-108-RU..>	2016-01-03 15:44	-	
Auto_user_PRO-113-20..>	2016-01-03 09:24	-	
Auto_user_PRO-114-RU..>	2016-01-03 03:42	-	
Auto_user_PRO-115-15..>	2016-01-02 18:27	-	
Auto_user_PRO-116-15..>	2016-01-02 14:09	-	
Auto_user_PRO-119-20..>	2016-01-02 12:28	-	

Index of /run_results/Auto_user_Proton-178-RUN041_MITO_2EPASSAGE_273_433

Name	Last modified	Size	Description
Parent Directory		-	
BlanKet/	2016-01-06 23:45	-	
Niourk/	2016-01-14 23:12	-	
log/	2016-01-06 15:58	-	
out/	2016-01-06 16:26	-	

Diagnostique NGS – Klauncher

```
#***** Barcode Loop *****#
for path_bam in `ls $path_run_data/*_rawlib.bam`
  do
    bam_name=$(basename $path_bam)
    barcode=$(echo $bam_name | sed 's/_rawlib.bam//g')

    # Check if barcode KIWI results already present
    if [ ! -f $path_run_share_out/Kiwi_$barcode.html ] || !(grep "SUCCESS" $path_run_share_out/Kiwi_$barcode.html > /dev/null)
        then
            #***** Launch Kiwi *****#
            python $kiwi -a ${1} -b $barcode | tee $path_run_share_out/Kiwi_$barcode.out
            # Convert program execution output to html
            cat $path_run_share_out/Kiwi_$barcode.out | $ansi2html --bg=dark > $path_run_share_out/Kiwi_$barcode.html
            # Delete temp out
            rm -f $path_run_share_out/Kiwi_$barcode.out

            #***** RSYNC Niourk_DB *****#
            rsync -rtv $path_NiourkDB/ $path_NiourkDB_diagNGSserver/ > /dev/null

            #***** Launch TVC (on both Kiwi BAMs) *****#
            if [ ! -f $path_run_tvc/$barcode.vcf ] ; then tvc -r $path_ref -b $path_run_kiwi/$barcode"_Kiwi.bam" -t $path_dist_design -O $path_run_tvc -o $barcode.
vcf --parameters-file $path_tvc_parameters_json -e /results/plugins/variantCaller/share/TVC/sse/motifset.txt >>/home/ionadmin/results/Auto_user_Proton-179-151214TS
CCUSTOMPANEL4POOLHiq_278_435/log/TVC.log 2>&1 ; fi
            if [ ! -f $path_run_tvc/$barcode"_pseudo.vcf".vcf ] ; then tvc -r $path_ref -b $path_run_kiwi/$barcode"_Kiwi_pseudo.bam" -t $path_dist_design -O $path_
run_tvc -o $barcode"_pseudo.vcf" --parameters-file $path_tvc_parameters_json -e /results/plugins/variantCaller/share/TVC/sse/motifset.txt >>/home/ionadmin/results/
Auto_user_Proton-179-151214TSCCUSTOMPANEL4POOLHiq_278_435/log/TVC.log 2>&1 ; fi

            #***** Transfer TVC vcf and BAMs to diagNGSserver *****#
            rsync -rtv $path_run_results_dir/ $path_run_diagNGSserver/ > /dev/null

            #***** Launch Niourk on diagNGSserver *****#
            ssh $diagNGSserver 'python /home/scripts/Niourk_v1_7.py -a ${1} -b $barcode'
    fi
done


#***** Niourk add frequency & Niourk Excel on diagNGSserver *****#
# Check all is finish
ssh $diagNGSserver 'python /home/scripts/niourk_add_frequency.py -a ${1}'
ssh $diagNGSserver 'python /home/scripts/niourk_to_excel.py -a ${1}'

#***** Launch BlanKet on diagNGSserver *****#
ssh $diagNGSserver 'python /home/scripts/BlanKet.py -a ${1}'

#***** Copy results file to run_results share folder *****#

#***** Delete temporary run folder *****#
rm -rf $path_tmp
```