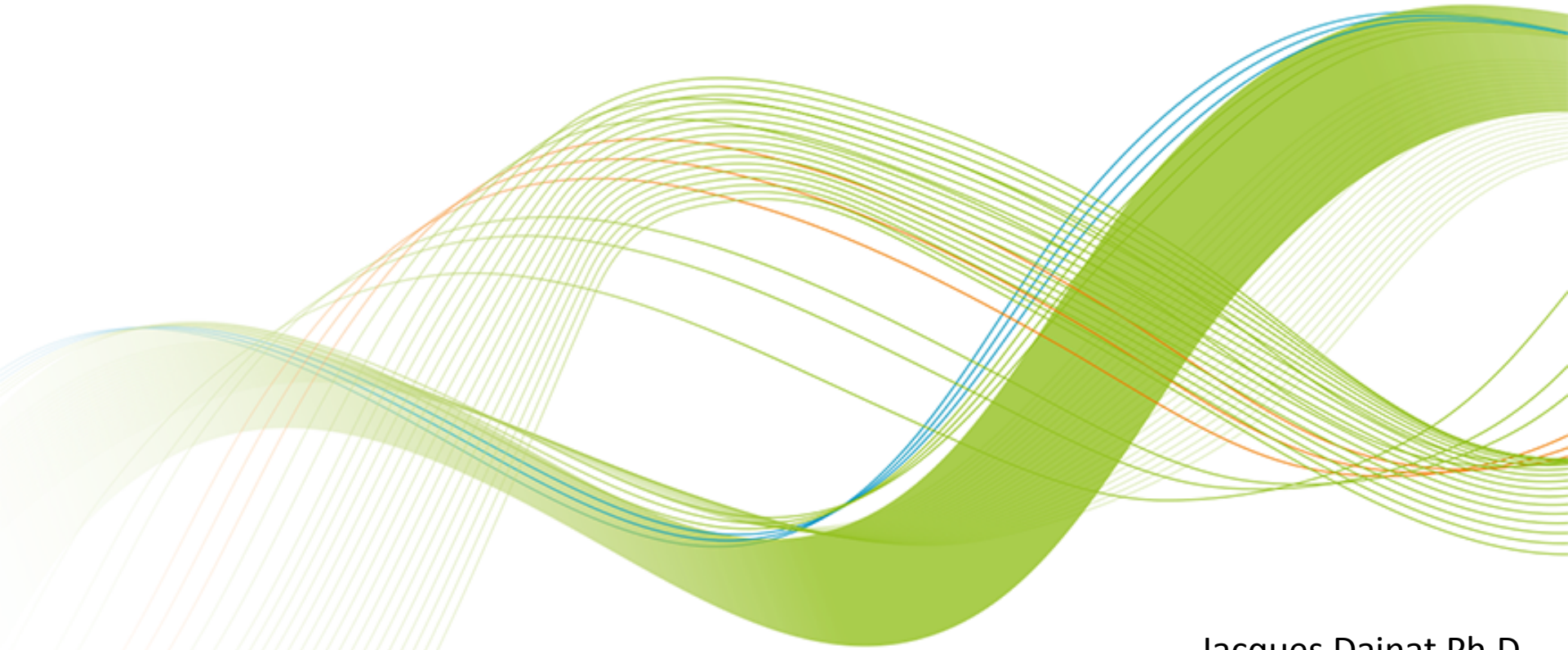
SciLifeLab

NBIS
NATIONAL BIOINFORMATICS INFRASTRUCTURE SWEDEN

A tool to check the library type of RNA-seq data



Jacques Dainat Ph.D.

There are numerous library preparation protocols for RNA-seq that result in sequencing reads with different characteristics.

- For example, reads can be single end (only one side of a fragment is recorded as a read) or paired-end (reads are generated from both ends of a fragment).

- Further, the sequencing reads themselves may be unstranded or strand-specific.

- Finally, paired-end protocols will have a specified relative orientation.

Transcriptome assembly tools need that kind of information to perform better.

e.g. mapper: Hisat2

`--rna-strandness <string>`

Specify strand-specific information: the default is unstranded.
For single-end reads, use F or R. 'F' means a read corresponds to a transcript. 'R' means a read corresponds to the reverse complemented counterpart of a transcript. For paired-end reads, use either FR or RF.
With this option being used, every read alignment will have an XS attribute tag: '+' means a read belongs to a transcript on '+' strand of genome. '-' means a read belongs to a transcript on '-' strand of genome.

(TopHat has a similar option, --library-type option, where fr-firststrand corresponds to R and RF; fr-secondstrand corresponds to F and FR.)

`--fr/--rf/--ff`

The upstream/downstream mate orientations for a valid paired-end alignment against the forward reference strand. E.g., if `--fr` is specified and there is a candidate paired-end alignment where mate 1 appears upstream of the reverse complement of mate 2 and the fragment length constraints (`-I` and `-X`) are met, that alignment is valid. Also, if mate 2 appears upstream of the reverse complement of mate 1 and all other constraints are met, that too is valid. `--rf` likewise requires that an upstream mate1 be reverse-complemented and a downstream mate2 be forward-oriented. `--ff` requires both an upstream mate 1 and a downstream mate 2 to be forward-oriented. Default: `--fr` (appropriate for Illumina's Paired-end Sequencing Assay).

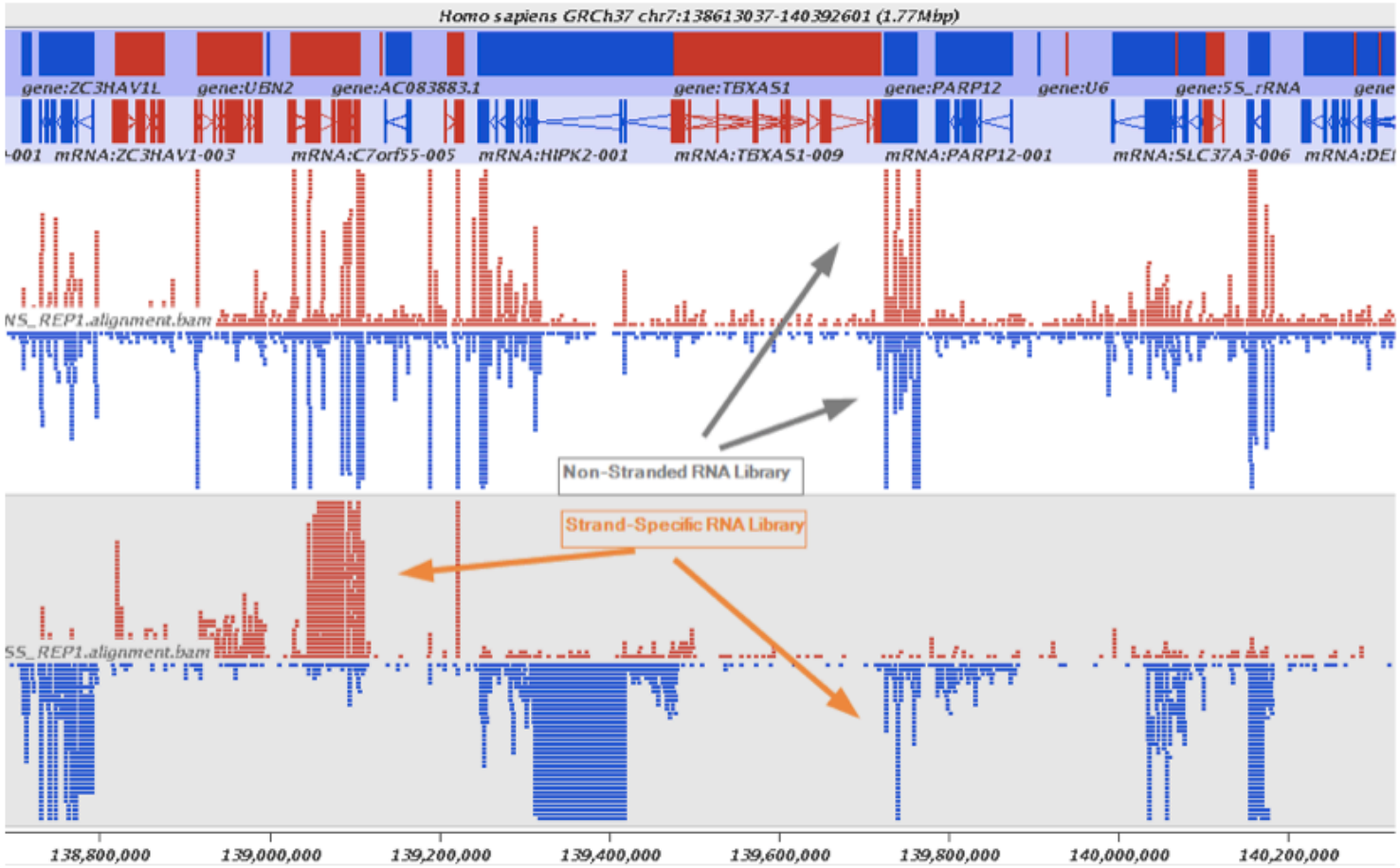
e.g. guided-assembler: Stringtie

`--rf`	Assumes a stranded library fr-firststrand.
`--fr`	Assumes a stranded library fr-secondstrand.

e.g. guided-assembler: Trinity

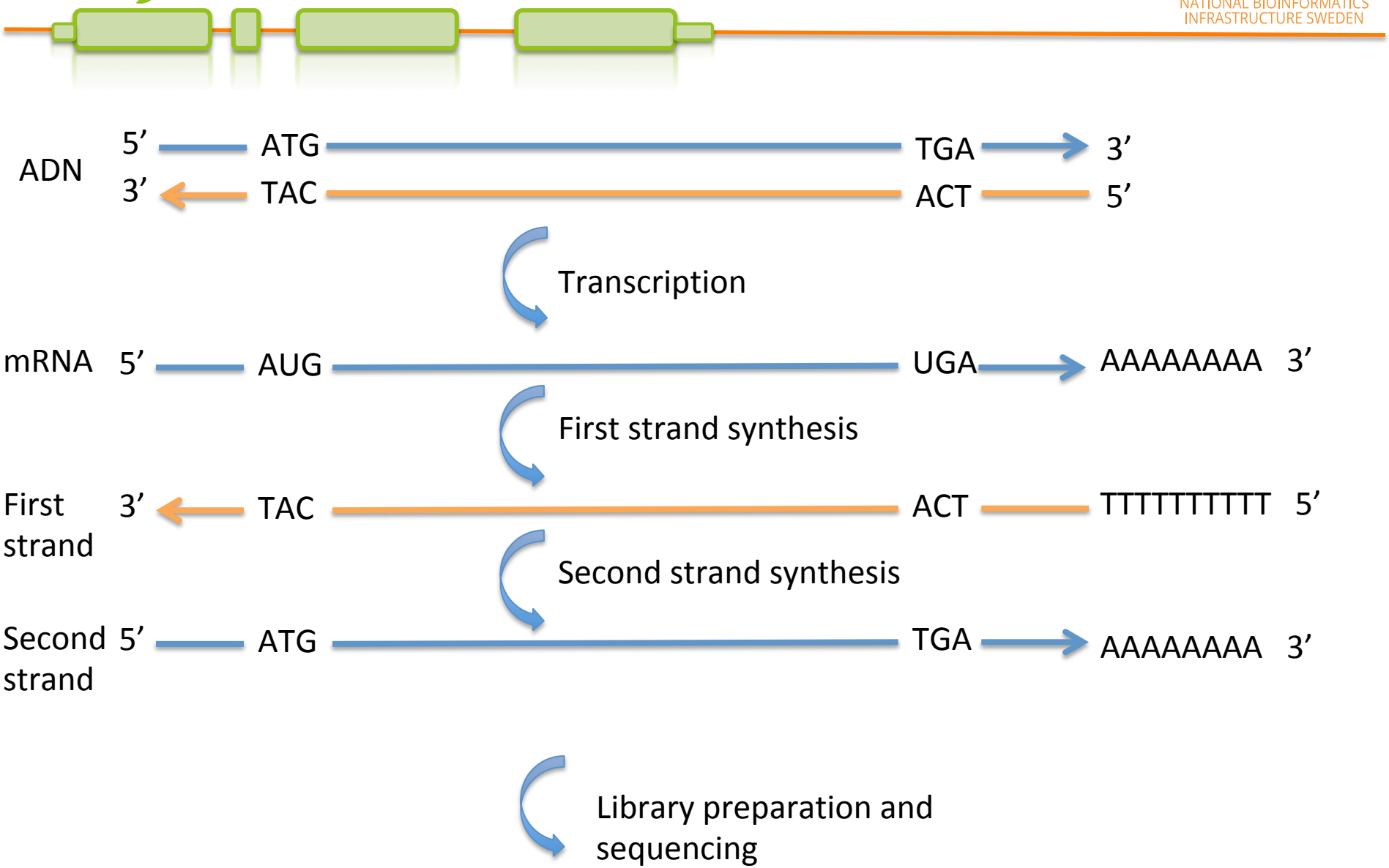
```
#  --SS_lib_type <string>          :Strand-specific RNA-Seq read orientation.
#                                   if paired: RF or FR,
#                                   if single: F or R.   (dUTP method = RF)
```

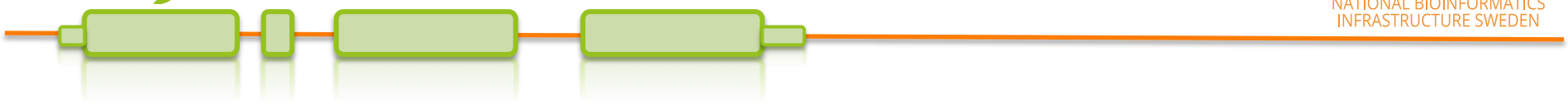
The implication of stranded RNAseq is that you can distinguish whether the reads are derived from forward- or reverse-encoded transcripts:



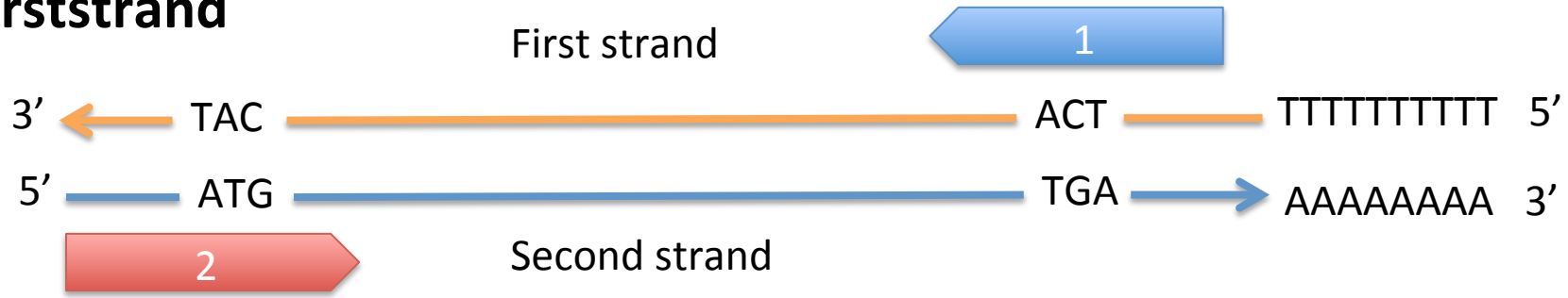
Stranded RNAseq data look like this

This example contrasts unstranded and stranded RNAseq experiments. Red transcripts are from + strand and blue are from - strand. In stranded example reads are clearly stratified between the two strands. A small number of reads from opposite strand may represent anti-sense transcription. The image from GATC Biotech.

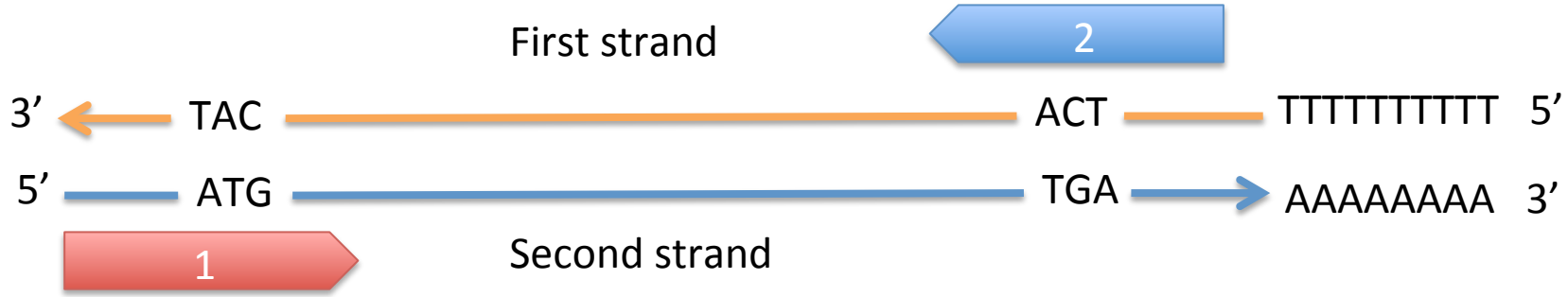

ADN
5’
ATG
TGA
3’
3’
TAC
ACT
5’
Transcription
mRNA
5’
AUG
UGA
AAAAAAAA
3’
First strand synthesis
First strand
3’
TAC
ACT
TTTTTTTTTT
5’
Second strand synthesis
Second strand
5’
ATG
TGA
AAAAAAAA
3’
Library preparation and sequencing



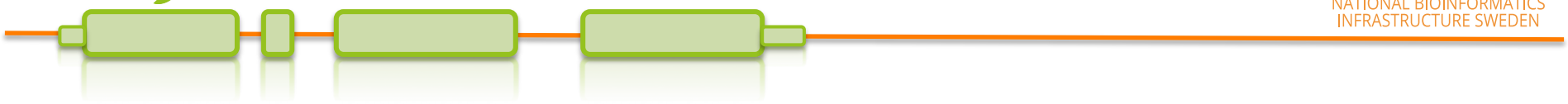
fr-firststrand



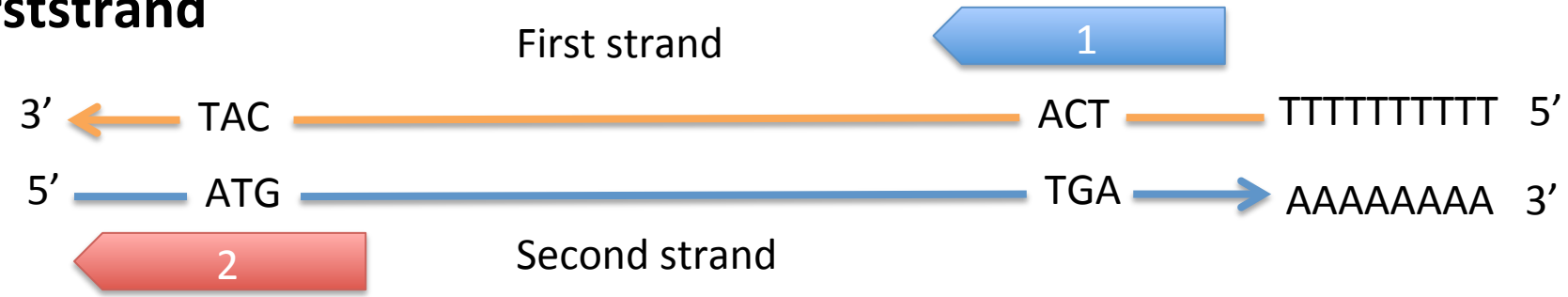
fr-secondstrand



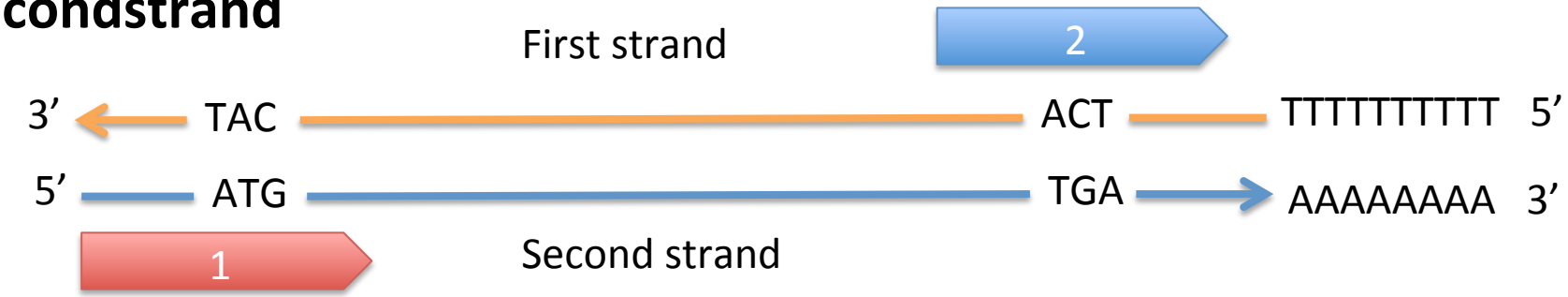
Mix of both = **fr-unstranded**



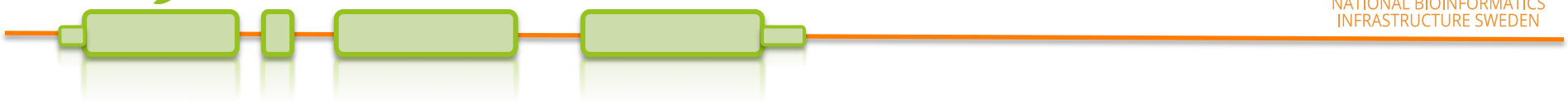
ff-firststrand



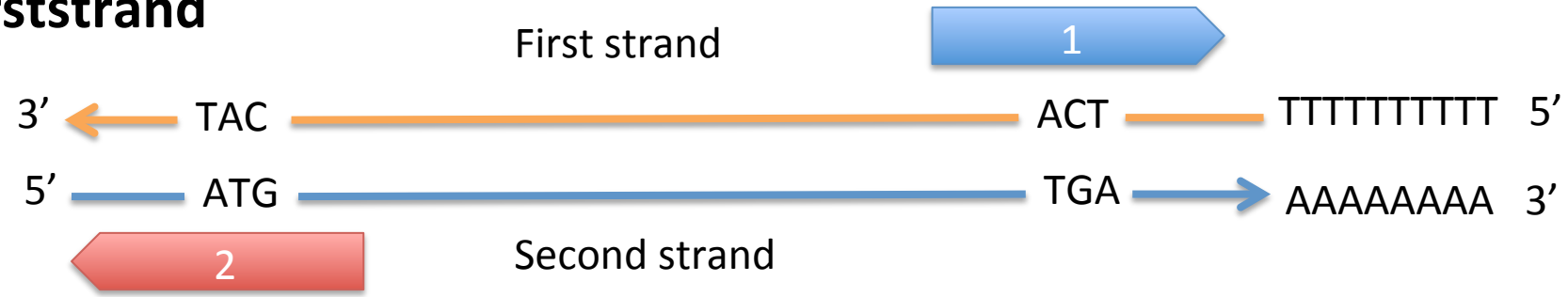
ff-secondstrand



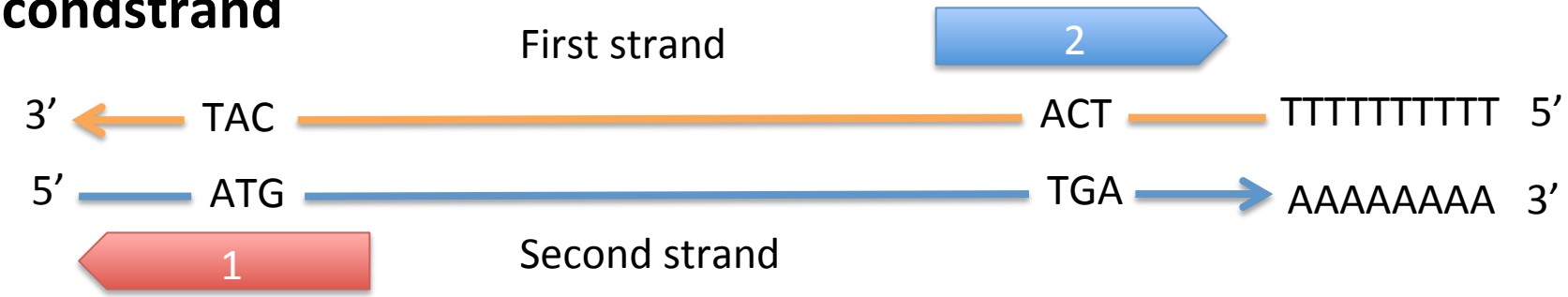
Mix of both = **ff-unstranded**



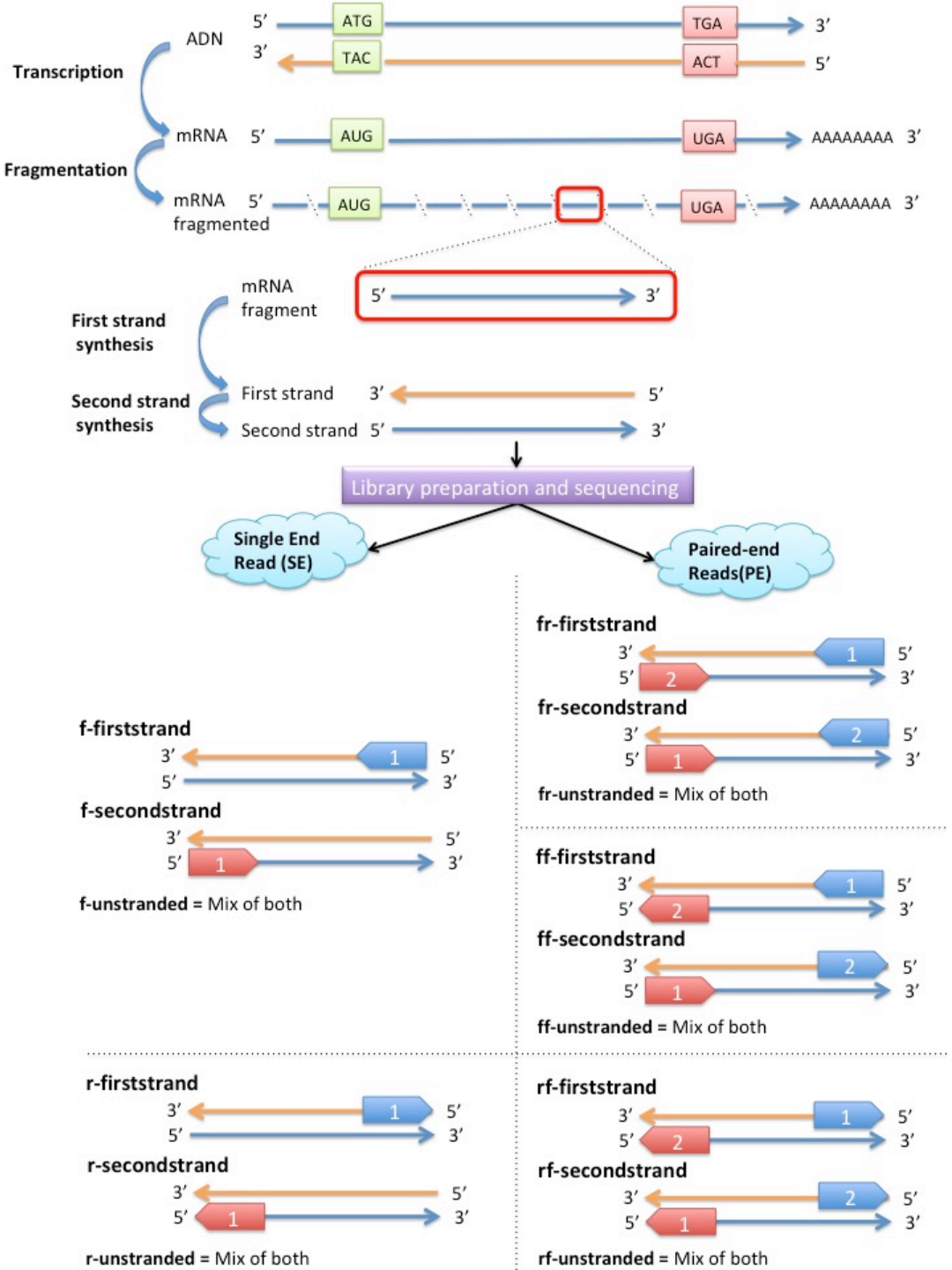
rf-firststrand



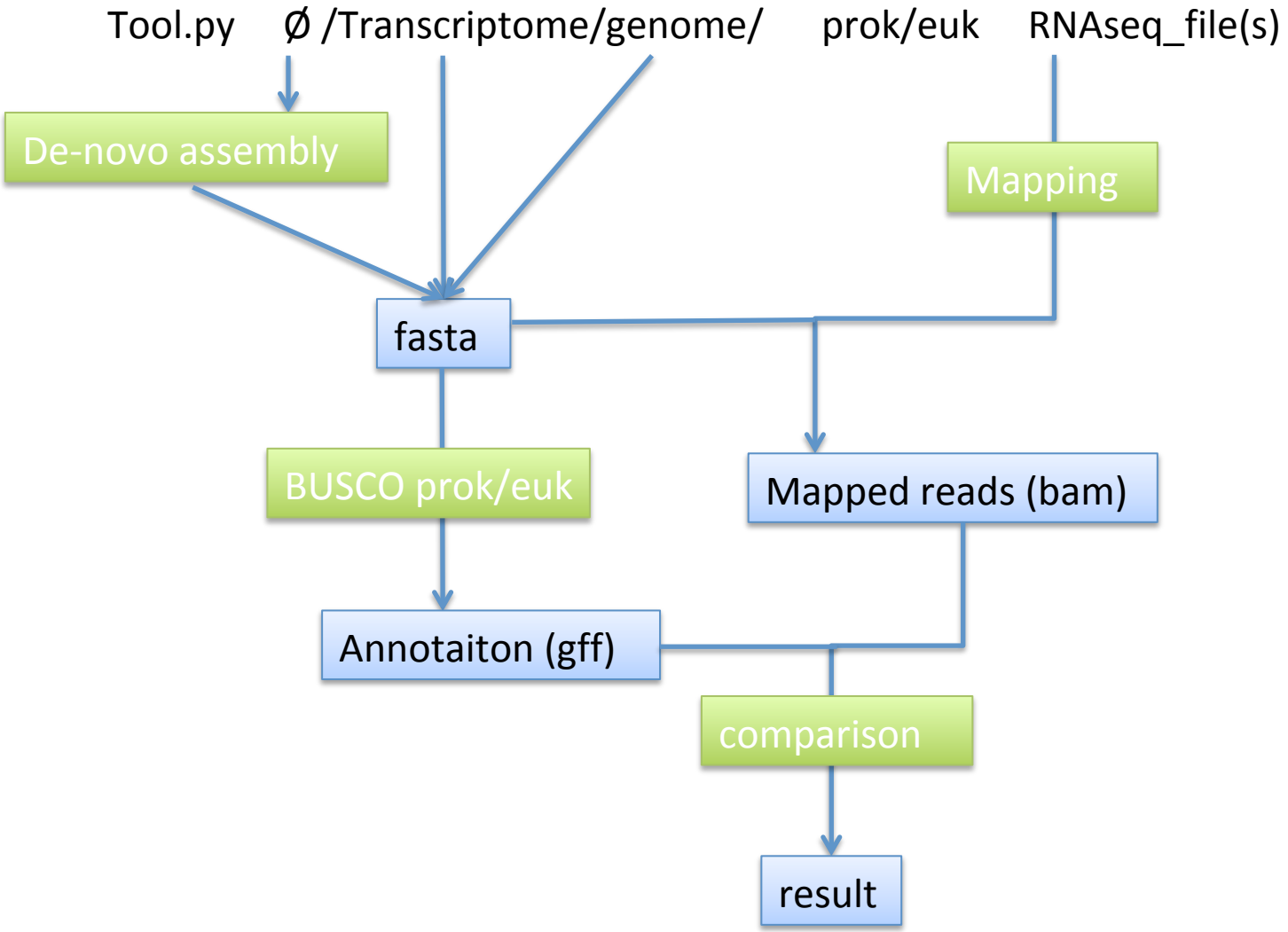
rf-secondstrand



Mix of both = **rf-unstranded**


ADN
5′
ATG
TGA
3′
3′
TAC
ACT
5′
Transcription
mRNA
5′
AUG
UGA
AAAAAAAA
3′
Fragmentation
mRNA fragmented
5′
AUG
UGA
AAAAAAAA
3′
mRNA fragment
5′
3′
First strand synthesis
Second strand synthesis
First strand
3′
5′
Second strand
5′
3′
Library preparation and sequencing
Single End Read (SE)
Paired-end Reads(PE)
f-firststrand
3′
1
5′
5′
3′
f-secondstrand
3′
5′
5′
1
3′
f-unstranded = Mix of both
fr-firststrand
3′
1
5′
5′
2
3′
fr-secondstrand
3′
2
5′
5′
1
3′
fr-unstranded = Mix of both
ff-firststrand
3′
1
5′
5′
2
3′
ff-secondstrand
3′
2
5′
5′
1
3′
ff-unstranded = Mix of both
r-firststrand
3′
1
5′
5′
3′
r-secondstrand
3′
5′
5′
1
3′
r-unstranded = Mix of both
rf-firststrand
3′
1
5′
5′
2
3′
rf-secondstrand
3′
2
5′
5′
1
3′
rf-unstranded = Mix of both

Workflow


Tool.py
Ø /Transcriptome/genome/
prok/euk
RNAseq_file(s)
De-novo assembly
Mapping
fasta
BUSCO prok/euk
Mapped reads (bam)
Annotaiton (gff)
comparison
result

- Read pairs in the --fr orientation are produced using the Illumina paired end protocol. Read pairs in the --rf orientation are produced using the Illumina mate-pair protocol. Read pairs in the --ff orientation are produced in using the SOLiD mate-pair protocol.