# NextFlow Introduction

*Maxime Borry, Quentin Letourneur*

*06/09/2018*

## Introduction

The goal of this exercise is to recreate the pipeline showned in Figure 1 represented by its DAG

## Channel factory

You may want to use the have a look at the `fromFilePairs` methods to create the channel when working with paired end data

## Mapping

The mapping command to use is the following:

```
bowtie2 -q -1 reads_1.fastq -2 reads_2.fastq -x index_prefix -S output.sam -p nb_cpus --very-sensitive-l
```

## Samtools view

The goal is to convert the `sam` file into a compressed binary `bam` file using `samtools view`

The samtools command to use is the following:

```
samtools view -S -@ nb_cpus -b -o output.bam input.sam
```

## Samtools sort

The goal is to sort the reads mapped on the reference genome by position

The samtools command to use us the following:

```
samtools sort -@ nb_cpus -o sorted_output.bam input.bam
```

## Bedtools

The goal is to compute the a position specific coverage of the reference genome provided with the aligned reads

The bedtools command to use is the following:

```
genomeCoverageBed -ibam sorted_input.bam > output.gcbout
```

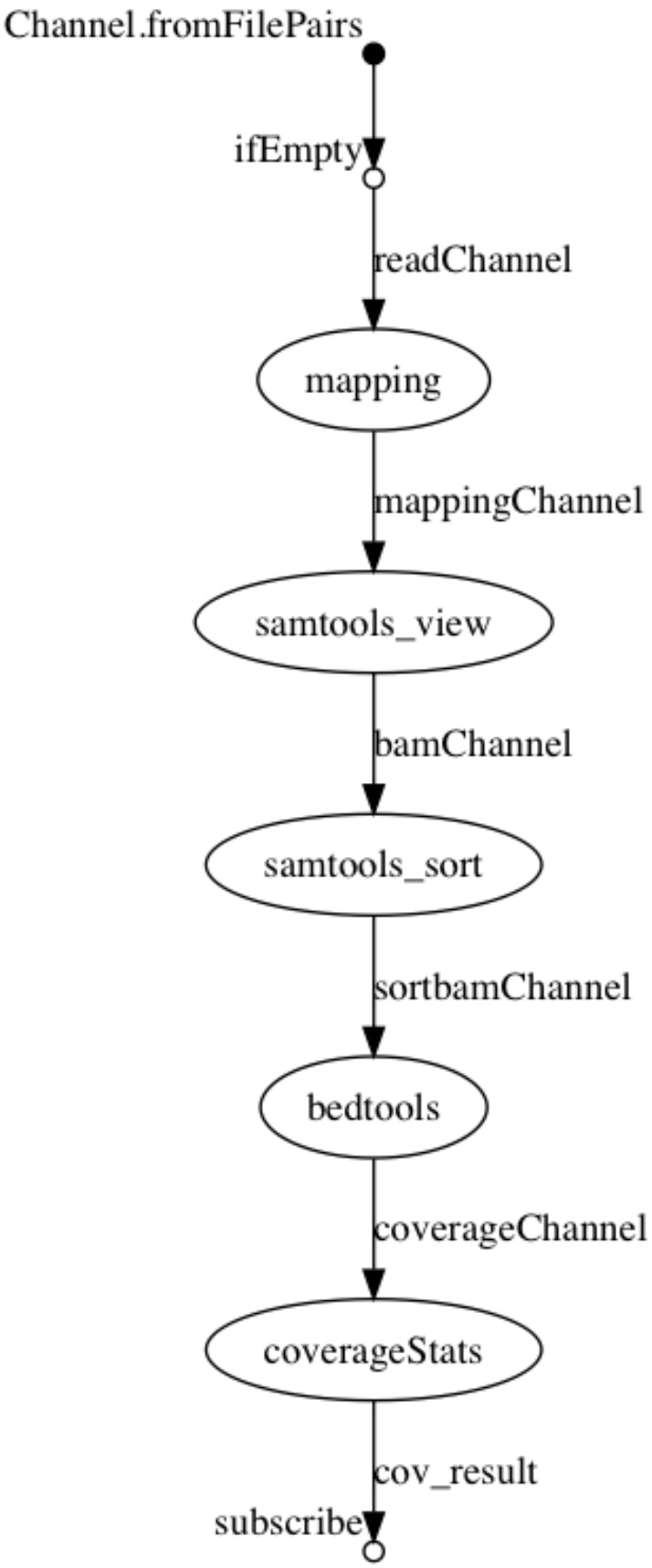


Figure 1: Directed Acyclic Graph (DAG) of the desired pipeline. Ellipses represent processes, arrows represents channels