https://training.nextflow.io/basic_training/intro/#your-first-script
https://www.youtube.com/watch?v=nzR8DRq13nE

2.4.1 Nextflow code

Info

Click the ⊕ icons in the code for explanations.

nf-training/hello.nf

```
1  #!/usr/bin/env nextflow
2
3  params.greeting = 'Hello world!'
4  greeting_ch = Channel.of(params.greeting)
5
6  process SPLITLETTERS {
7      input:
8      val x
9
10     output:
11     path 'chunk_*'
12
13     script:
14     """
15     printf '$x' | split -b 6 - chunk_
16     """
17 }
18
19 process CONVERTTOUPPER {
20     input:
21     path y
22
23     output:
24     stdout
25
26     script:
27     """
28     cat $y | tr '[a-z]' '[A-Z]'
29     """
30 }
31
32 workflow {
33     letters_ch = SPLITLETTERS(greeting_ch)
34     results_ch = CONVERTTOUPPER(letters_ch.flatten())
35     results_ch.view { it }
36 }
```

#shebang declares nextflow as interpreter

#initialize channel

#val or path

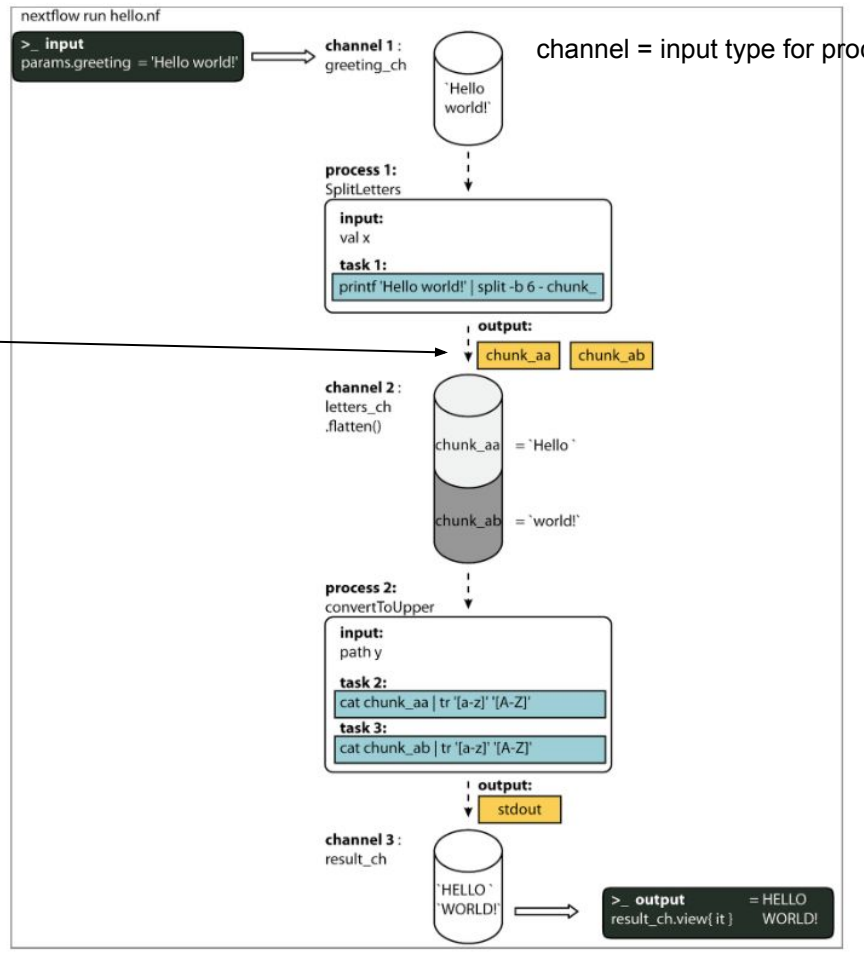
#output file with prefix 'chunk_'

#expect standard output

#store outputs in channels

2.7 In DAG-like format

To better understand how Nextflow is dealing with the data in this workflow, below is a DAG-like figure to visualize all the inputs, outputs, channels and processes:



channel = input type for processes

Specify fq's for RNA seq

```
nf-training > script3.nf
1  #!/usr/bin/env nextflow
2
3  /*
4   * pipeline input parameters
5   */
6  params.reads = "$projectDir/data/ggal/gut_{1,2}.fq"
7  params.transcriptome_file = "$projectDir/data/ggal/transcriptome.fa"
8  params.multiqc = "$projectDir/multiqc"
9  params.outdir = "results"
10
11 log.info """\
12     R N A S E Q - N F   P I P E L I N E
13     ===================================
14     transcriptome: ${params.transcriptome_file}
15     reads        : ${params.reads}
16     outdir       : ${params.outdir}
17     """
18     .stripIndent()
19
20 read_pairs_ch = Channel.fromFilePairs(params.reads)
21 read_pairs_ch.view()
```

Gut files _1 | _2

```
PROBLEMS 59  OUTPUT  DEBUG CONSOLE  TERMINAL
gitpod /workspace/gitpod/nf-training (master) $ nextflow run script3.nf
                                      $ nextflow run script3.nf --reads "data/ggal/gut_{1,2}.fq"

Launching `script3.nf` [determined_gates] DSL2 - revision: e0e4df9340

R N A S E Q - N F   P I P E L I N E
===================================
transcriptome: /workspace/gitpod/nf-training/data/ggal/transcriptome.fa
reads        : /workspace/gitpod/nf-training/data/ggal/gut_{1,2}.fq
outdir       : results

[gut, [/workspace/gitpod/nf-training/data/ggal/gut_1.fq, /workspace/gitpod/nf-training/data/ggal/gut_2.fq]]
```

Base name file (str): gut
file (path) pairs: gut1fq, gut2fq

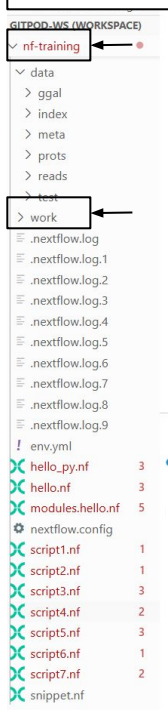
```
gitpod /workspace/gitpod/nf-training (master) $ nextflow run script3.nf --reads "data/ggal/*_{1,2}.fq"
```

Wildcard fq's for RNAseq

```
R N A S E Q - N F   P I P E L I N E
===================================
transcriptome: /workspace/gitpod/nf-training/data/ggal/transcriptome.fa
reads        : data/ggal/*_{1,2}.fq
outdir       : results

[gut, [/workspace/gitpod/nf-training/data/ggal/gut_1.fq, /workspace/gitpod/nf-training/data/ggal/gut_2.fq]]
[liver, [/workspace/gitpod/nf-training/data/ggal/liver_1.fq, /workspace/gitpod/nf-training/data/ggal/liver_2.fq]]
[lung, [/workspace/gitpod/nf-training/data/ggal/lung_1.fq, /workspace/gitpod/nf-training/data/ggal/lung_2.fq]]
```

Folder formatting for RNA seq



workdir

Outdir, executed processes

```
PROBLEMS 59  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS  SPELL CHECKER 41
gitpod /workspace/gitpod/nf-training (master) $ ls
data      hello.nf     modules.hello.nf  script1.nf  script3.nf  script5.nf  script7.nf  work
env.yml   hello_py.nf  nextflow.config   script2.nf  script4.nf  script6.nf  snippet.nf
```