Analysis protocol for GENE-SWitCH capture Hi-C data (pig and chicken)
Version of December 12th, 2022

Pig and chicken GENE-SWitCH capture Hi-C data (*in situ Hi-C followed by capture enrichment and high throughput sequencing*) were processed using the nf-core hic pipeline (https://github.com/nf-core/hic) version 1.3.0.

For pig and using a slurm cluster, we performed the following steps:

1) **Input data**

From hereafter the species name, either <sus_scrofa> for pig or <gallus_gallus> for chicken, will be refered to as `$species`.

Input data are stored in a <data.$species> directory

a) Get raw sequencing reads in fastq.gz format from the ENA
- *Pig:* Accessions PRJEB44198 and PRJEB44486
- *Chicken:* Accessions PRJEB53986 and PRJEB54773

b) Get the genomic sequences ready (here downloaded from ENSEMBL build 102)
- Pig: sus_scrofa.fa (Sscrofa11.1)
- Chicken: gallus_gallus (GRCg6a)

c) Get the corresponding bowtie2 indices, named after the genome file $species.fa.*.bt2 (Bowtie2 version: 2.3.5.1) with the command

```
bowtie --index $species.fa
bowtie2-build $species.fa $species.fa
```

d) Get a file with the chromosomes and contigs' lengths using for instance the fastalength tool from the Exonerate package version 2.2.0

```
fastalength $species.fa | awk '{print $2"\t"$1}' > chrom.len
```

2) **Launch the nf-core hic on the GENE-SWitCH capture Hi-C reads**

a) make a config file `capturehic.conf`

```
singularity {
   enabled = true
 }
process {
 withName:bowtie2_end_to_end {
    cpus = 24
    memory = 64.GB
```

```
    time = 36.h
  }
  withName:trim_reads {
    memory = 32.GB
    time = 72.h
  }
  withName:bowtie2_on_trimmed_reads {
    cpus = 16
    memory = 64.GB
    time = 36.h
  }
  withName:bowtie2_merge_mapping_steps {
    cpus = 8
    time = 36.h
  }
  withName:get_valid_interaction {
    memory = 64.GB
    time = 36.h
  }
  withName:remove_duplicates {
    cpus = 32
    memory = 160.GB
    time = 36.h
  }
  withName:tads_hicexplorer {
    memory = 64.GB
    time = 36.h
  }
  withName:tads_insulation {
    memory = 32.GB
    time = 36.h
  }
  withName:dist_decay {
    memory = 64.GB
    time = 36.h
  }
  withName:compartment_calling {
    memory = 32.GB
    time = 36.h
  }
  withName:combine_mates {
    memory = 16.GB
  }
  withName:merge_stats {
```

```
    time = 8.h
  }
  withName:multiqc {
    memory = 8.GB
    time = 3.h
  }
  withName:build_contact_maps {
    memory = 32.GB
    time = 6.h
  }
}
```

b) *make a bash script called* `capturehic.$species.sh` *to be launched (here on a slurm cluster) with the following content:*

```
#!/bin/bash
mkdir -p capturehic.$species
cd capturehic.$species

module load bioinfo/nfcore-Nextflow-v21.10.6
nextflow run nf-core/hic \
 -revision    1.3.0 \
 -profile     genotoul \
 -resume               \
 -c capturehic.conf \
 -work-dir capturehic.$species/work \
 --fasta data.$species/$species.fa \
 --bwt2_index data.$species \
 --outdir capturehic.$species \
 --input 'data.$species/*_R{1,2}.fastq.gz'\
 --bwt2_opts_end2end '-5 5 -3 95 --very-sensitive -L 20 --score-min
L,-0.6,-0.3 --end-to-end --reorder' \
 --bwt2_opts_trimmed '--very-sensitive -L 20 --score-min L,-0.6,-0.3
--end-to-end --reorder' \
 --min_mapq 10 \
 --restriction_site '^GATC,G^ANTC' \
 --ligation_site 'GATCGATC,GANTGATC,GANTANTC,GATCANTC' \
 --chromosome_size data.$species/chrom.len \
 --min_insert_size 10 \
 --max_insert_size 1000 \
 --bin_size                  '500000,10000,5000,1000' \
 --ice_max_iter     100 \
 --ice_filter_low_count_perc  0.02 \
 --ice_filter_high_count_perc 0 \
```

```
 --ice_eps     0.1 \
 --res_compartments           '500000,200000' \
 --tads_caller                'insulation,hicexplorer' \
 --res_tads                   '100000,50000' \
 --save_interaction_bam \
 --skip_balancing \
 --skip_tads
```

c) launch it on a slurm cluster with the command

```
sbatch capturehic.$species.sh
```

3) Submitted analysis files

a) Hi-C interaction matrices (contact maps)

Those are contact matrices in .cool format, one per fastq.gz read pair, that were located in the `capturehic.$species/contact_maps/raw/cool` directory