

TOGGLE

Tools for Generic NGS analysis

A framework to quickly build pipelines
and to perform large-scale NGS analysis

Christine Tranchant-Dubreuil
christine.tranchant@ird.fr

TOGGLE



- A toolbox to perform large-scale NGS analyses

19 modules, 88 functions
40 open-source tools



TOGGLE



- A toolbox to perform large-scale NGS analyses

19 modules, 88 functions
40 open-source tools



- **Version 2 published in BMC bioinformatics**

RESEARCH

TOGGLE: Toolbox for generic NGS analyses

Cecile Monat^{1*†}, Christine Tranchant-Dubreuil^{1†^}, Ayité Kougbéadjó², Cédric Farcy², Mawussé Agbessi¹, Maryline Summo² and Francois Sabot^{1^}

Data preprocessing

Fastqc, Cutadapt
FastxTrimmer
Stack process_radstats

Structural Variations

MindTheGap,
BreakDancer, Pindel

RNA-seq Assembly

Trinity
TGI-CL

ReadCount

Htseq-count

TOGGLE



Mapping

Bwa aln, sampe/ samse
Bwa mem
Tophat2

SNP calling/ filtering

SAMtools, GATK, VarScan, SNPEff

SAM/BAM management

picardTools, SAMtools, GATK

Data preprocessing

Fastqc, Cutadapt
FastxTrimmer
Stack process_radstats

Structural Variations

MindTheGap,
BreakDancer, Pindel

RNA-seq Assembly

TransAbyss, Trinity
TGI-CL

ReadCount

Htseq-count

TOGGLE



Mapping

Bwa aln, sampe/ samse
Bwa mem
Tophat2

SNP calling/ filtering

SAMtools, GATK, VarScan, SNPEff

SAM/BAM management

picardTools, SAMtools, GATK



<https://github.com/SouthGreenPlatform/TOGGLE>

TOGGLE



Version 2



Version 3

*From hard-coded pipelines
To a bioinformatic pipeline framework*

TOGGLE



Version 2



Version 3

*From hard-coded pipelines
To a bioinformatic pipeline framework*

Biologists create their own pipeline through an easy and user-friendly approach



How to perform an analysis with TOGGLE^{Le} ?

A command-line based pipeline framework



A single command line

```
toggleGenerator.pl -d DIR-c FILE -o DIR
```


What does **TOGGLE** need to run ?

- An input directory (with fastq, sam/bam, vcf files)
- The name of output directory used to store the data generated by the analyses
- A unique and simple configuration file to design the pipeline and define software parameters.
- Optional arguments : reference file, annotation...

\$order

1=fastqc

2=cutadapt

3=bwa mem

4=picardToolsSortSam

5=samToolsView

1000=gatkHaplotypeCaller

1001=gatkVariantFiltration

\$cutadapt

-q 30

-m 35

\$bwa mem

-n 5

...

\$sge

-q bioinfo.q

-b Y

\$order

1=fastqc

2=cutadapt

3=bwa mem

4=picardToolsSortSam

5=samToolsView

1000=gatkHaplotypeCaller

1001=gatkVariantFiltration

Create your own workflow

- The workflow order
- The list of softwares to run

One line = the step followed by the software's name

Create your own workflow

\$order

1=fastqc

2=cutadapt

3=bwa mem

4=picardToolsSortSam

5=samToolsView

1000=gatkHaplotypeCaller

1001=gatkVariantFiltration

Step number < 1000

Step number >= 1000

\$order

1=fastqc

2=cutadapt

3=bwa mem

4=picardToolsSortSam

5=samToolsView

1000=gatkHaplotypeCaller

1001=gatkVariantFiltration

Create your own workflow

Step number < 1000

Parallel analysis by sample

\$order

1=fastqc

2=cutadapt

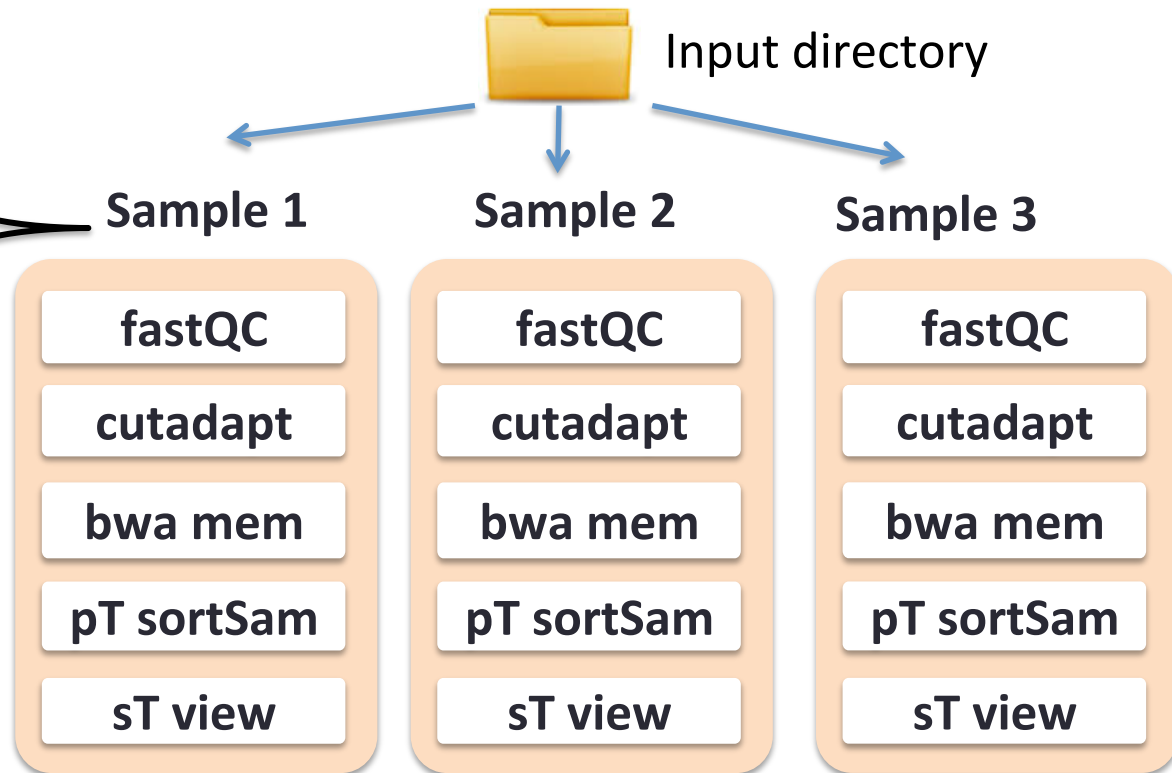
3=bwa mem

4=picardToolsSortSam

5=samToolsView

1000=gatkHaplotypeCaller

1001=gatkVariantFiltration



\$order

1=fastqc

2=cutadapt

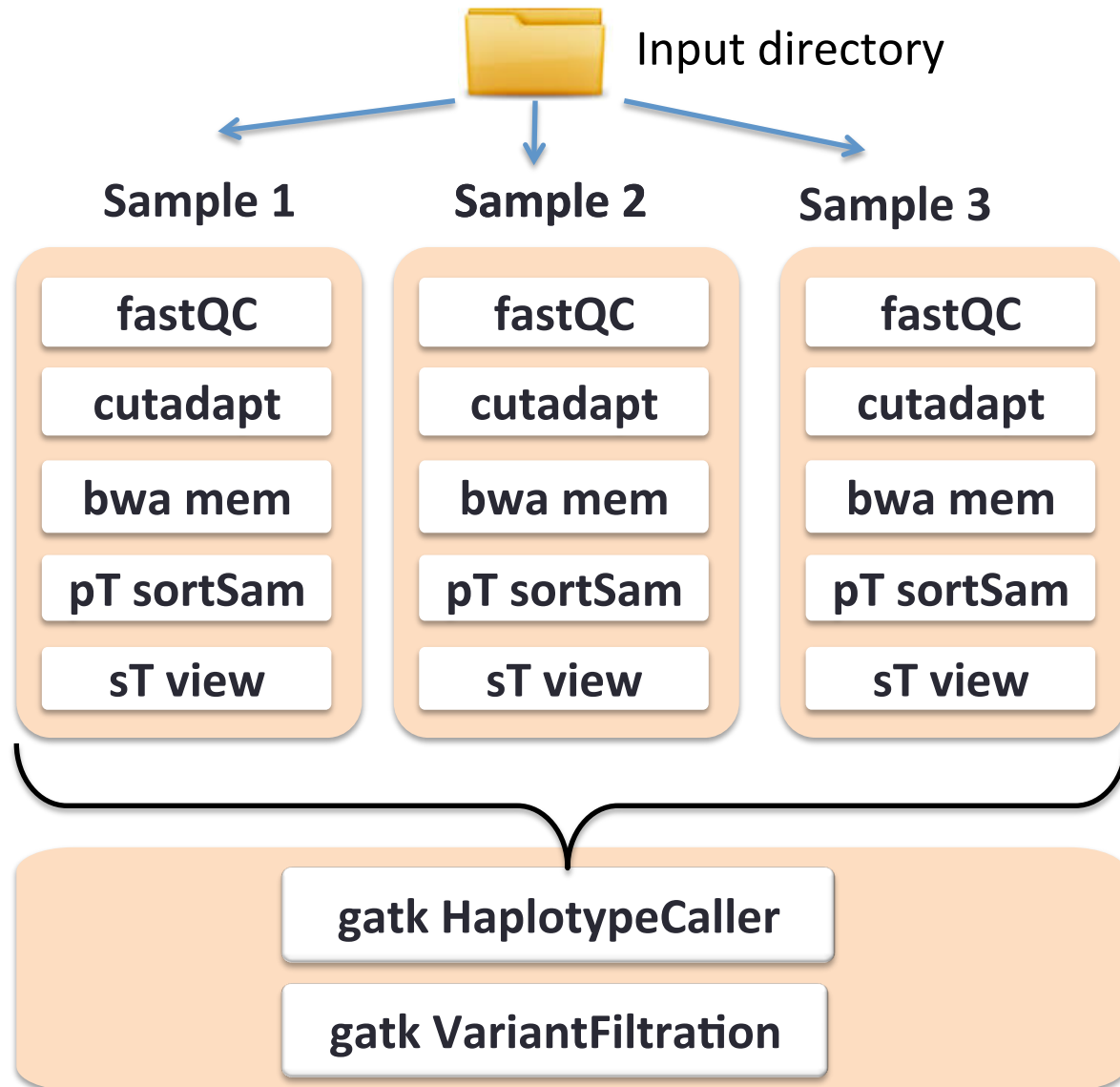
3=bwa mem

4=picardToolsSortSam

5=samToolsView

1000=gatkHaplotypeCaller

1001=gatkVariantFiltration

Step number ≥ 1000 **Global analysis
(all samples)**

\$order

1=fastqc

2=cutadapt

3=bwa mem

4=picardToolsSortSam

5=samToolsView

1000=gatkHaplotypeCaller

1001=gatkVariantFiltration

\$cutadapt

-q 30

-m 35

\$bwa mem

-n 5

...

\$sge

-q bioinfo.q

-b Y

Software parameters

One tag per software (\$softwareName)
followed by the list of options

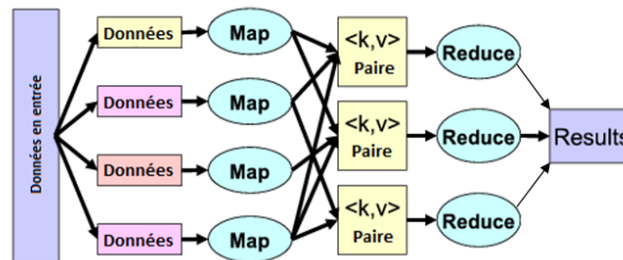
Demo

- New tools

VCFTools
Bcftools



- New data analysis : metagenomics, pacbio assembly
- New features : automatic PDF reports, non-sequential pipelines



TOGGLe 's team



UMR DIADE

Christine Tranchant-Dubreuil

François Sabot

Cécile Monat

Mawussé Agbessi

Souhila Amazougarene

Abdoulaye Diallo

Laura Helou

Ayité Kougbéadjó



UMR BGPI

Sébastien Ravel



Julie Orjuela-Bouniol



UMR AGAP

Cédric Farcy

Enrique Ortega-Abboud

Gautier Sarah

Maryline Summo



Thank you for your attention !



The screenshot shows the GitHub interface for the repository **SouthGreenPlatform / TOGGLE**. At the top, navigation links include Personal, Open source, Business, Explore, Pricing, Blog, and Support. A search bar and buttons for Sign in and Sign up are also present. The repository page shows 17 watches, 5 stars, and 3 forks. Below this, tabs for Code, Issues (1), Pull requests (0), Pulse, and Graphs are visible. The selected branch is **master**, and the file **TOGGLE / README.md** is open. A commit by **francoissabot** is shown, dated 389699d on 16 Mar. The README content is displayed below, featuring the title **TOGGLE : Toolbox for generic NGS analyses** and a large logo where the 'G' is orange. The text describes TOGGLE as a suite of 10 packages and 110 modules for managing NGS software pipelines. It also mentions that users can create their own pipelines through a user-friendly approach.

SouthGreenPlatform / **TOGGLE** Watch 17 Star 5 Fork 3

<> Code Issues 1 Pull requests 0 Pulse Graphs

Branch: master TOGGLE / README.md Find file Copy path

francoissabot Adjusting the manual, readme and version to the TOGGLE correct git 389699d on 16 Mar

9 contributors

Executable File 154 lines (104 sloc) 4.95 KB Raw Blame History

TOGGLE : Toolbox for generic NGS analyses

TOGGLE

TOGGLE (TOolbox for Generic nGs anaLYsEs) is a suite of 10 packages and more than 110 modules able to manage a large set of NGS softwares and utilities to easily design pipelines able to handle hundreds of samples. Moreover, TOGGLE offers an easy way to manipulate the various options of the different softwares through the pipelines in using a single basic configuration file, that can be changed for each assay without having to change the code itself.

Users can also create their own pipeline through an easy and user-friendly approach. The pipelines can starts from Fastq (plain or gunzipped), SAM, BAM or VCF (plain of gunzipped) files, with parallel and global analyses. Samples pipelines are provided for SNP discovery and RNAseq counts.



<https://github.com/SouthGreenPlatform/TOGGLE>