

Integrated Workflow for Automated Transcription Start Site Prediction in Prokaryotes

Jonas Ditz, Friederike Hanssen, Julian Späth

Data Management in Quantitative Biology
Dr. Sven Nahnsen
Sven Fillinger
University of Tübingen

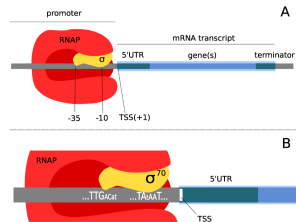
July 13, 2017

Overview

- 1 Introduction
- 2 Workflow
- 3 Outlook

Transcription Start Side

- Position of 5' UTR start
- Important for gene finding
- Predicted using enriched libraries and statistical models



Source: Data Management in Quantitative Biology,
Project 4 Sheet, summer term 2017

Workflow languages

- Automate processes
- Modular
- Open-Source

nextflow



Source: <https://www.nextflow.io/>
[https://bitbucket.org/snakemake/snakemake/wiki/](https://bitbucket.org/snakemake/snakemake/wiki/Documentation)
Documentation
<https://software.broadinstitute.org/wdl/>

Nextflow

- Docker/Singularity integration
- Cloud integration
- Extensive manual
- Several users (e.g. WHO)

International Agency
Research on Cancer



World Health
Organization

Docker

- Container management system
- Source on GitHub
- Images on DockerHub
- Exchangeable images



Source: www.docker.com

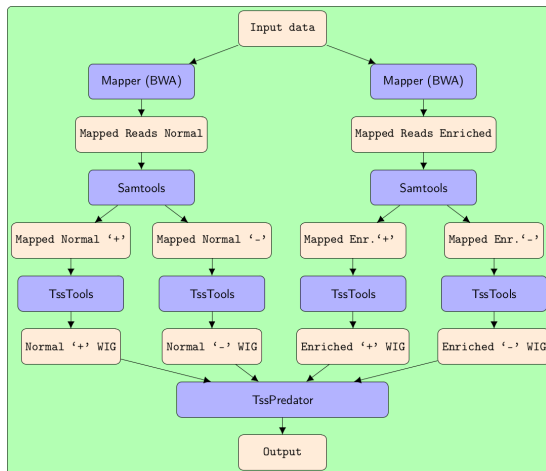
Overview

1 Introduction

2 Workflow

3 Outlook

Workflow Structure



Results

- Runs on all tested platforms (Mac, Linux Mint, Ubuntu (VM))
- Sequential and parallelized execution
- Highly modularized

Overview

1 Introduction

2 Workflow

3 Outlook

Future Improvements

- New Docker Modules
- Cloud Integration
- Investigate slow runtime of TssTools on Mac