Tomte: a pipeline for RNA-seq analysis in rare disease diagnostics

Lucía Peña-Pérez[1,2,3], Anders Jemt[2,4], Jesper Eisfeldt[2,3,4], Ramprasad Neethiraj[5], Esmee ten Berk de Boer[2,3,4], Mei Wu[6], Nicole Lesko[1], Anna Lindstrand[3], Henrik Stranneheim[1,2,4], Anna Wedell[1], Valtteri Wirta[2,4,5]

[1]Karolinska University Hospital, Centre for Inherited Metabolic Diseases, Stockholm, Sweden; [2]Science for Life Laboratory, Department of Microbiology, Tumour and Cell Biology, Karolinska Institutet, Stockholm, Sweden; [3]Karolinska Institutet, Department of Molecular Medicine and Surgery, Stockholm, Sweden; [4]Karolinska University Hospital, Genomics Medicine Center Karolinska, Stockholm, Sweden; [5]Science for Life Laboratory, School of Engineering Sciences in Chemistry, Biotechnology and Health, KTH Royal Institute of Technology, Stockholm, Sweden; [6]University of Washington, Department of Genome Sciences, Seattle, United States

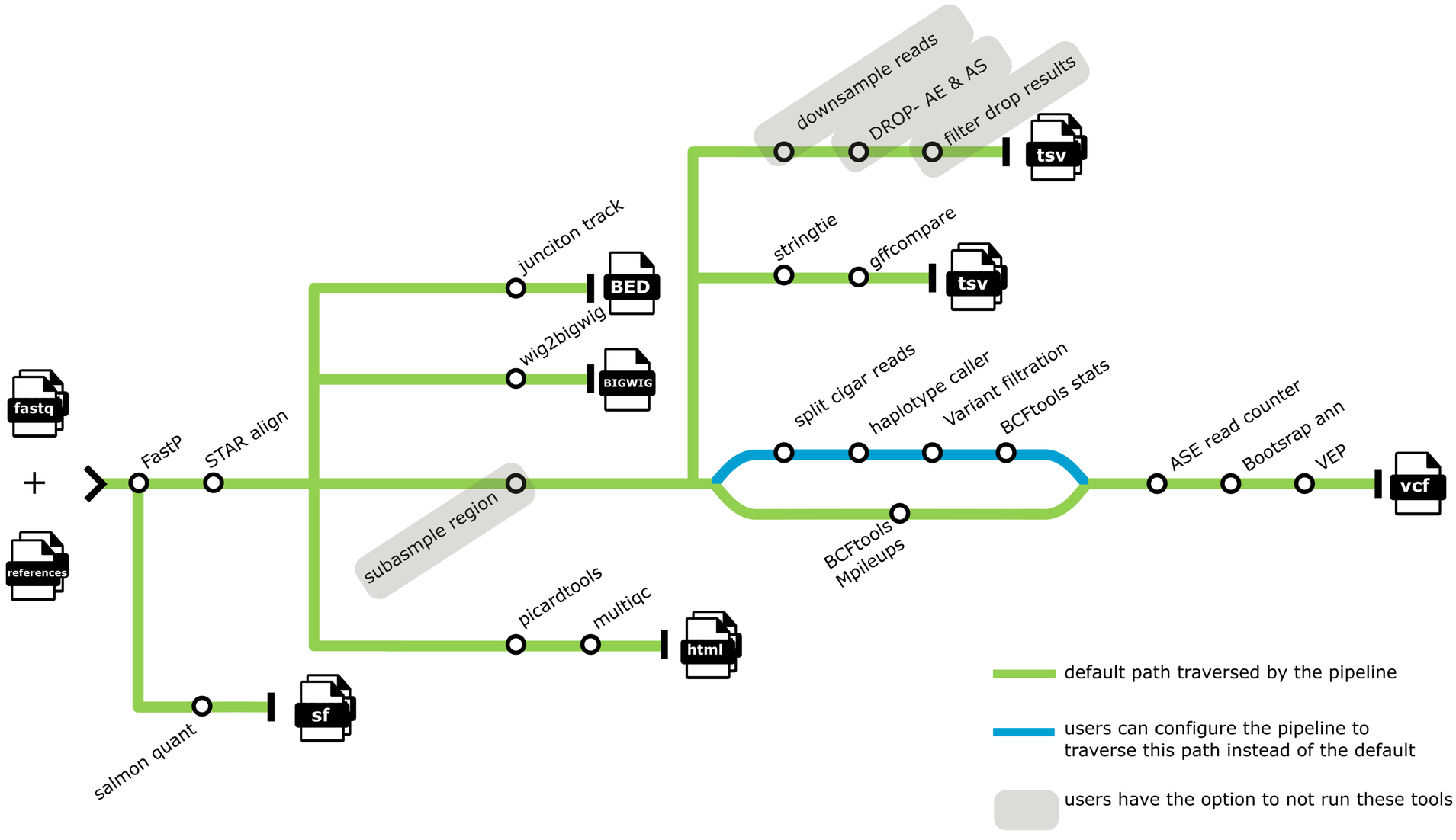
Background

- A rare disease is a condition affecting less than 1 in 2000 people
- They are rare individually but common as a whole, resulting in about 30 million people affected in the EU
- Most of these diseases have a clear genetic component but the cause of many of them remains unknown
- The diagnostic rate after performing WES or WGS is estimated at 30-50%

Motivation

- WGS analysis often yields many non-coding and deep intronic variants, the effect of which can be difficult to predicted reliably by in silico algorithms
- Whole transcriptome sequencing has shown to be a useful tool, particularly in seeing the effect of such variants in transcripts
- It allows for comprehensive detection of aberrant expression, aberrant splicing, and mono allelic expression in expressed genes

Workflow



Results & Conclusions

- Tomte is written in Nextflow DSL2 following nf-core recommendations allowing for reproducibility, portability, and continuous integration
- The pipeline will be used for investigating RNA-seq data from rare disease patient in the Stockholm healthcare region both in a research and clinical setting
- Providing a complete analysis including QCs, tracks, quantification, transcript assembly, and detection of aberrant expression, aberrant splicing, and mono-allelic expression

tomte

lucia.pena-perez@regionstockholm.se