nf-core/ airrflow

A pipeline to analyze Adaptive Immune Receptor Repertoire sequencing data with the Immcantation framework

Gisela Gabernet[1,*], Susanna Márquez[2,*], Alexander Peltzer[3], David Ladd[4], Robert Bjornson[5,6], Simon Heumos[1], Christoph Ruschil[7,8], Markus C. Kowarik[7,8], Steven H. Kleinstein[2,9,10], Sven Nahnsen[1,11], nf-core community.

[1]Quantitative Biology Center, Eberhard-Karls University of Tübingen, Tübingen, Germany. [2]Department of Pathology, Yale School of Medicine, New Haven, USA. [3]Böhringer Ingelheim Pharma GmbH & Co. KG, Biberach, Germany. [4]Biomedicine Discovery Institute, Monash University, Clayton, Australia. [5]Yale Center for Research Computing, New Haven, USA. [6]Department of Computer Science, Yale University, New Haven, USA. [7]Department of Neurology and Stroke, Center for Neurology, Eberhard-Karls University of Tübingen, Tübingen, Germany. [8]Hertie-Institute for Clinical Brain Research, Eberhard-Karls Universtiy of Tübingen, Tübingen, Germany. [9]Program in Computational Biology and Bioinformatics, Yale University, New Haven, USA. [10]Department of Immunobiology, Yale School of Medicine, New Haven, USA. [11]Biomedical Data Science, Department of Computational Science, Eberhard-Karls University of Tübingen, Germany. *These authors contributed equally to the work.

What are AIRRs?

The collection of B-cell receptors (BCR) and T-cell receptors (TCR) in an individual is referred to as the **Adaptive Immune Receptor Repertoire**. BCRs in the secreted form are also called immunoglobulins or antibodies.
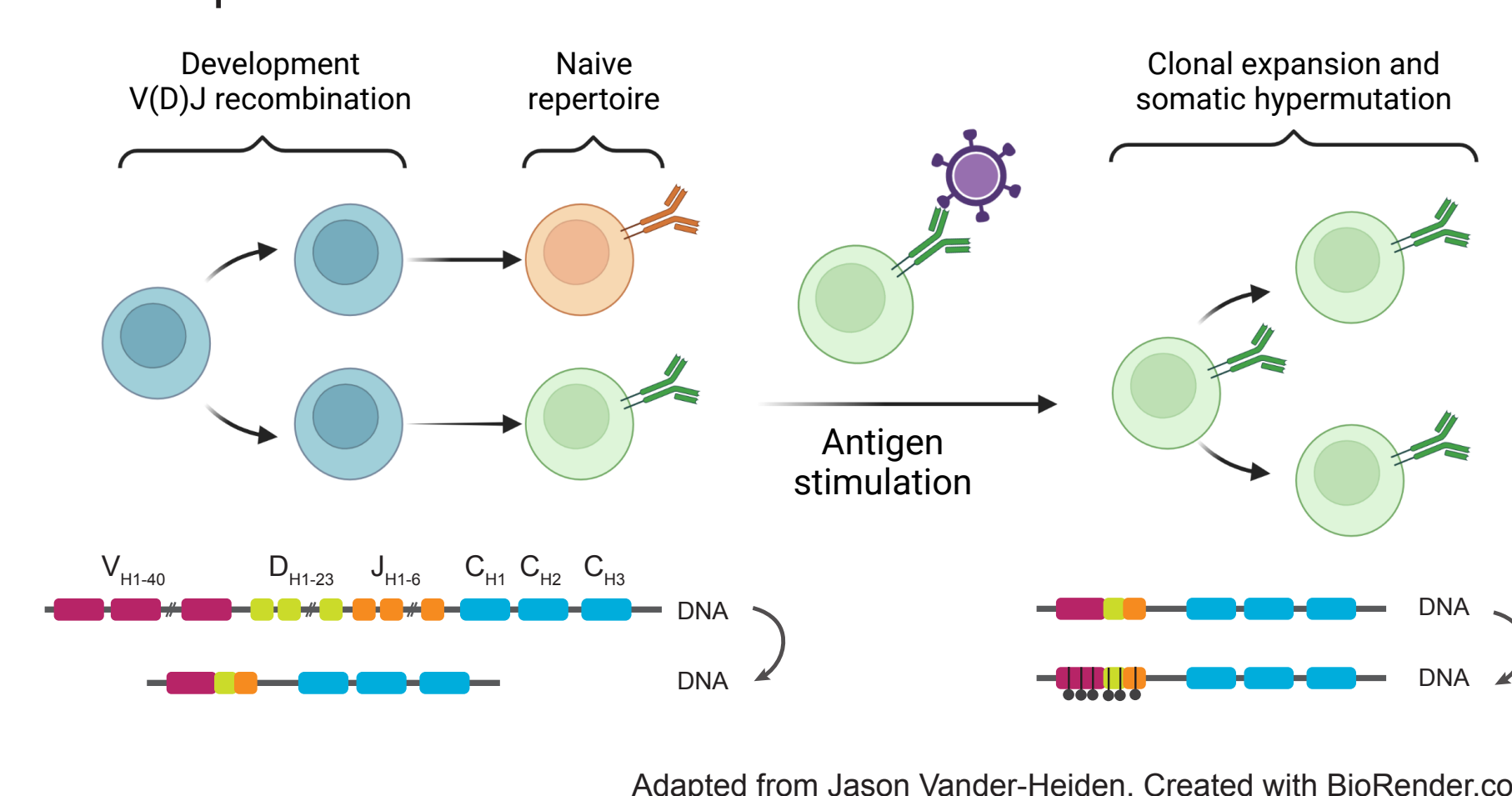
AIRR sequencing allows researchers to:

- Study the immune state of individuals
- Identify signatures of immune responses
- Detect immune-related diseases
- Guide vaccine development
- Guide immunotherapy

B-cell T-cell

10^{10-11} receptors in human body
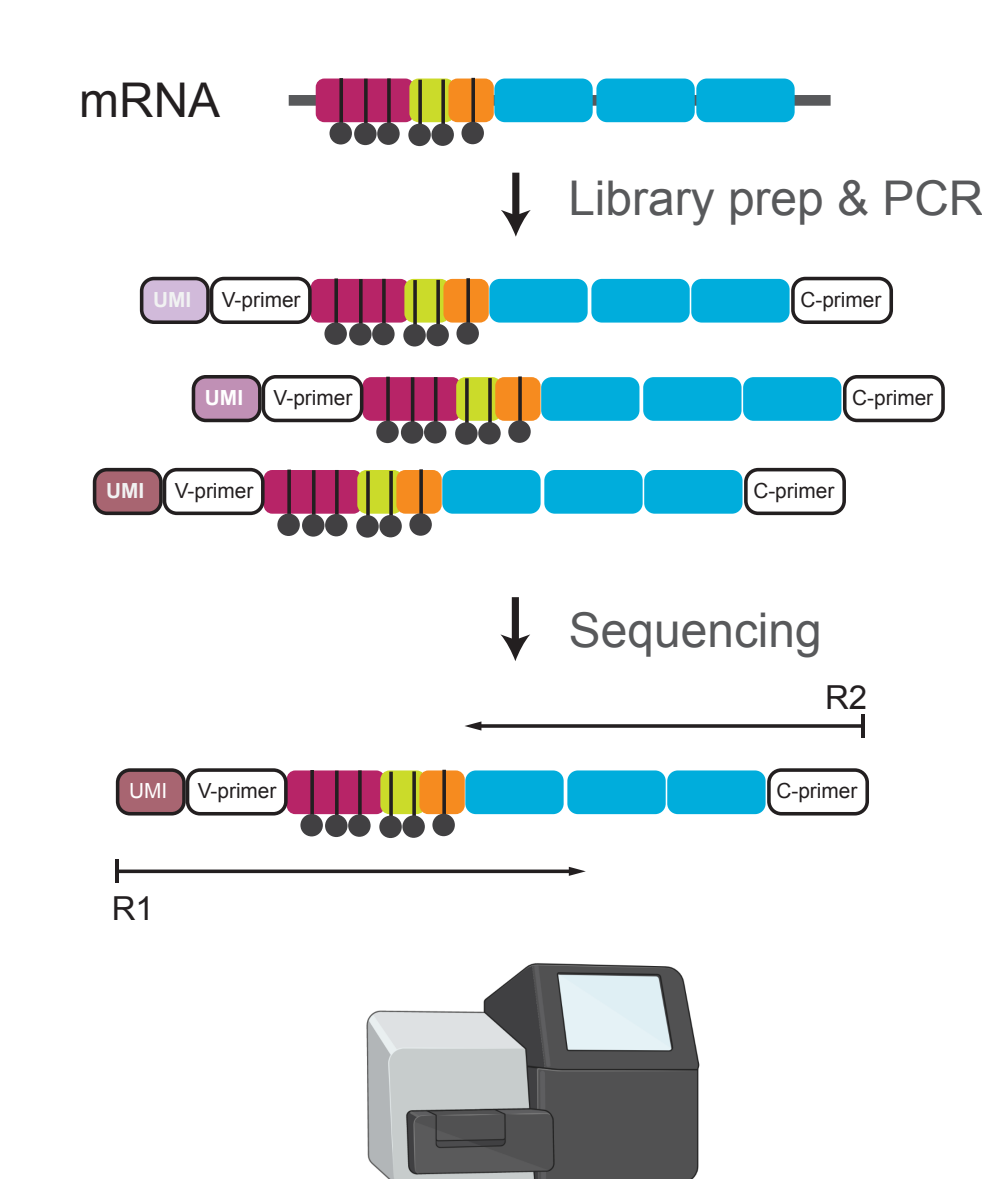
BCR / Ab

TCR

Created with BioRender.com
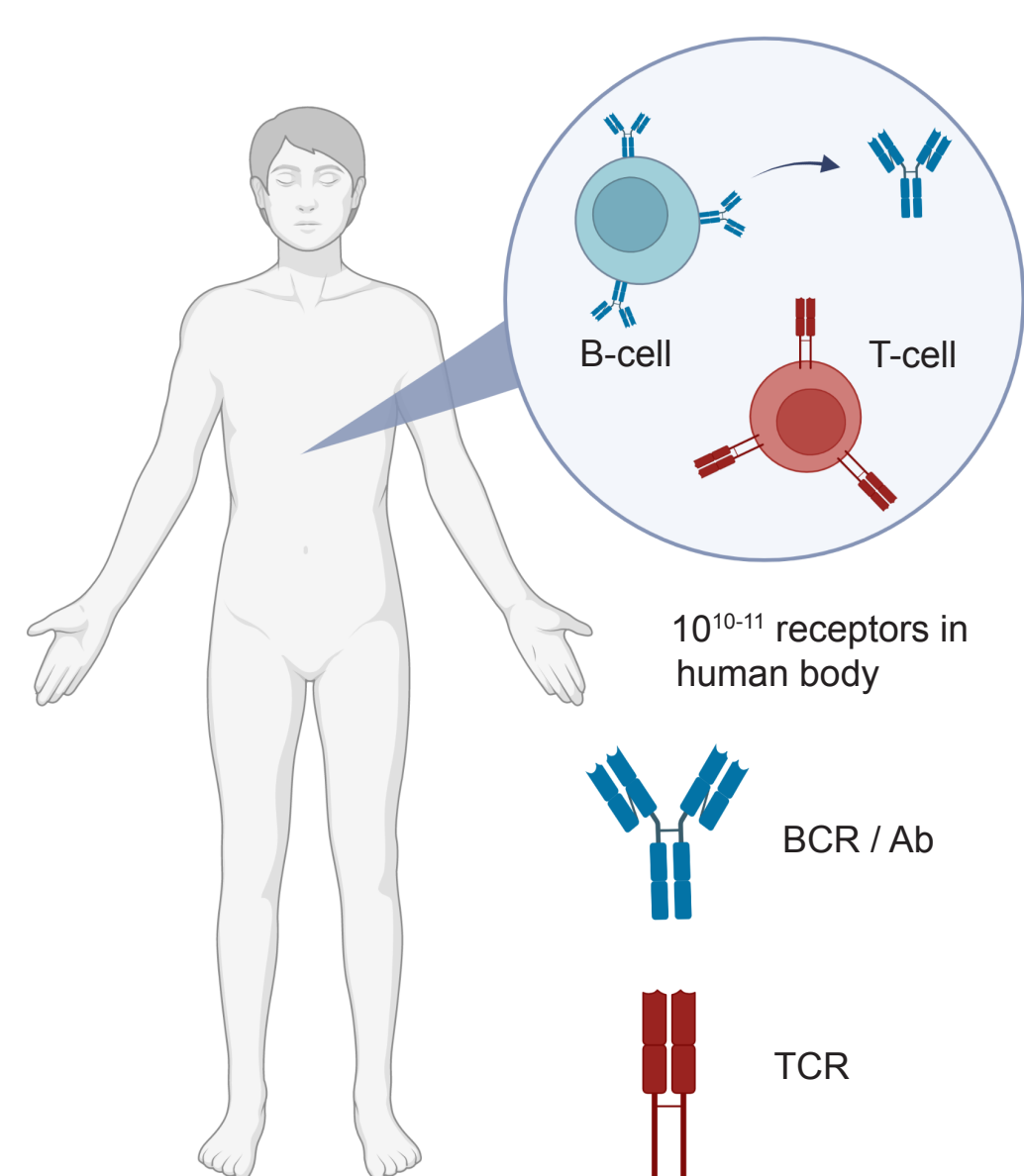
AIRR sequencing

The **AIRR diversity** is generated by **V(D)J somatic recombination** at the genomic level. In BCRs, additional **somatic hypermutation** occurrs after antigen stimulation. Due to these processes, the theoretical diversity of BCRs in humans is estimated at 10^{14} sequences.
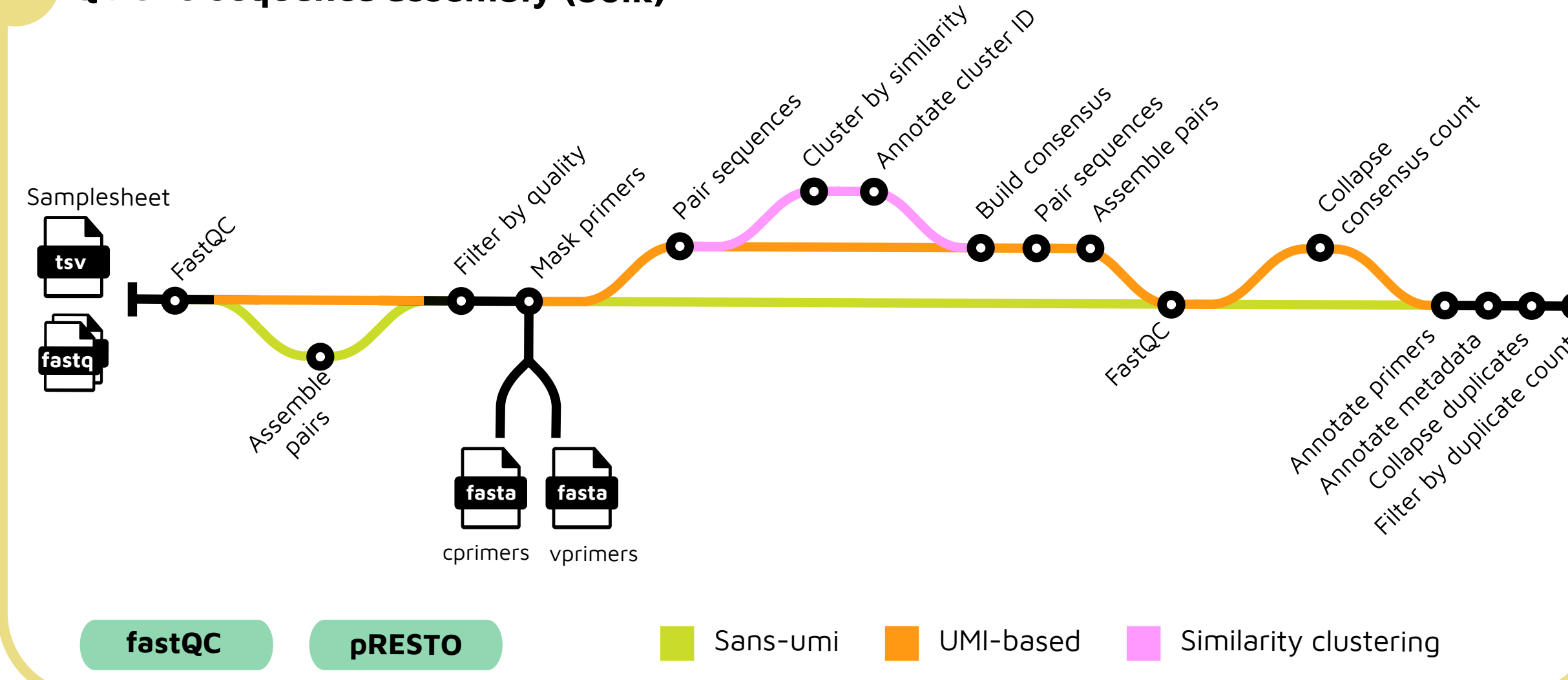
Development V(D)J recombination

Naive repertoire

Clonal expansion and somatic hypermutation

Antigen stimulation

V_{H1-40} D_{H1-23} J_{H1-6} C_{H1} C_{H2} C_{H3} DNA

Adapted from Jason Vander-Heiden, Created with BioRender.com

UMI-barcoded PCR sequencing

mRNA

↓ Library prep & PCR

UMI V-primer C-primer

↓ Sequencing

R2

R1

AIRR data analysis workflow

✓ **nf-core/airrflow** employs tools from the **Immcantation framework** to provide a start-to-end analysis of AIRR sequencing data.

✓ The workflow supports both **bulk** and **single-cell sequencing data**.

✓ nf-core/airrflow is written in **Nextflow**, fully **containerized**. It benefits from Nextflow's implicit parallelism for **high-throughput** analyses.

✓ Part of the **nf-core project**, it is **open source**, provides **stable releases**, extensive **documentation, CI testing**. Each release is tested on **AWS cloud** on full-size data.
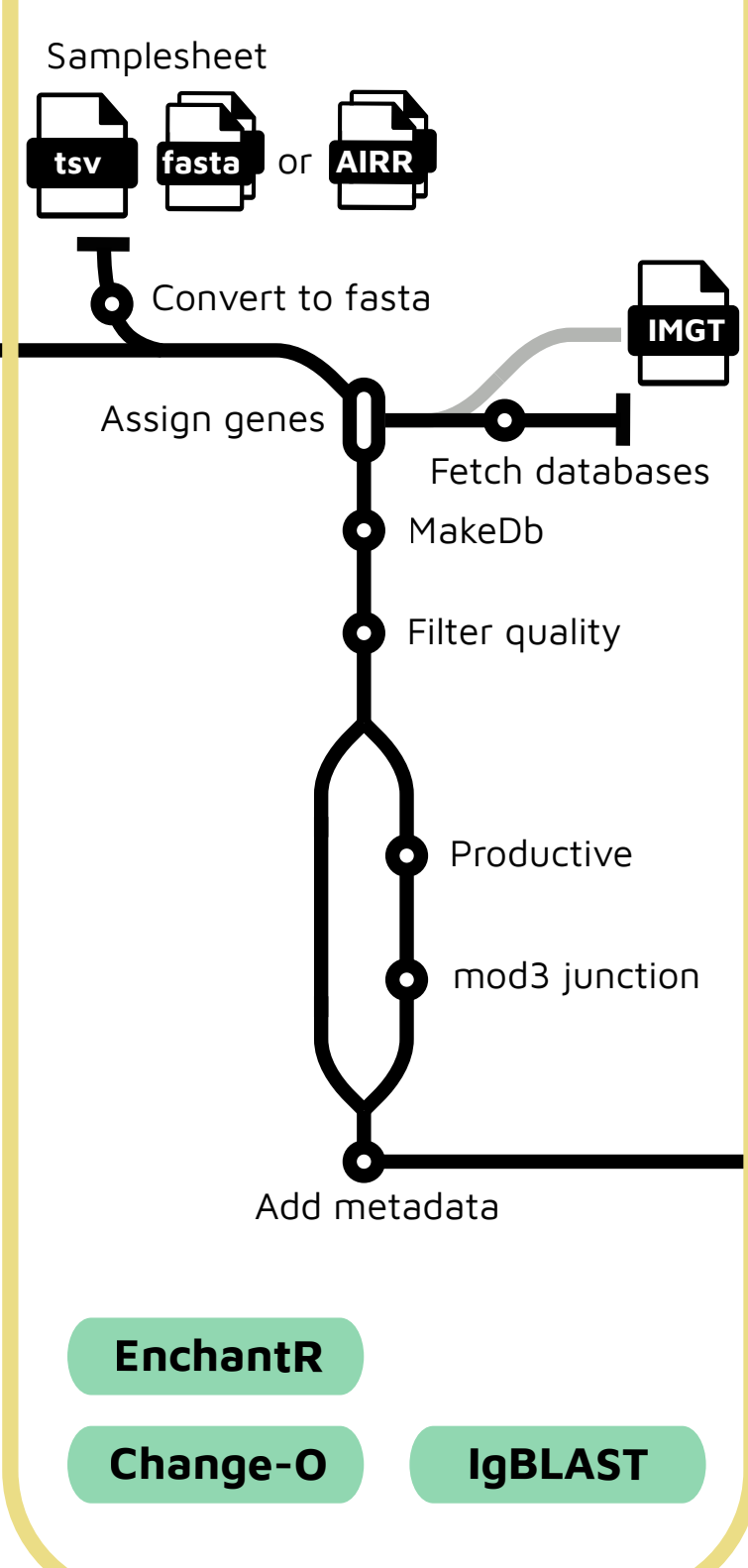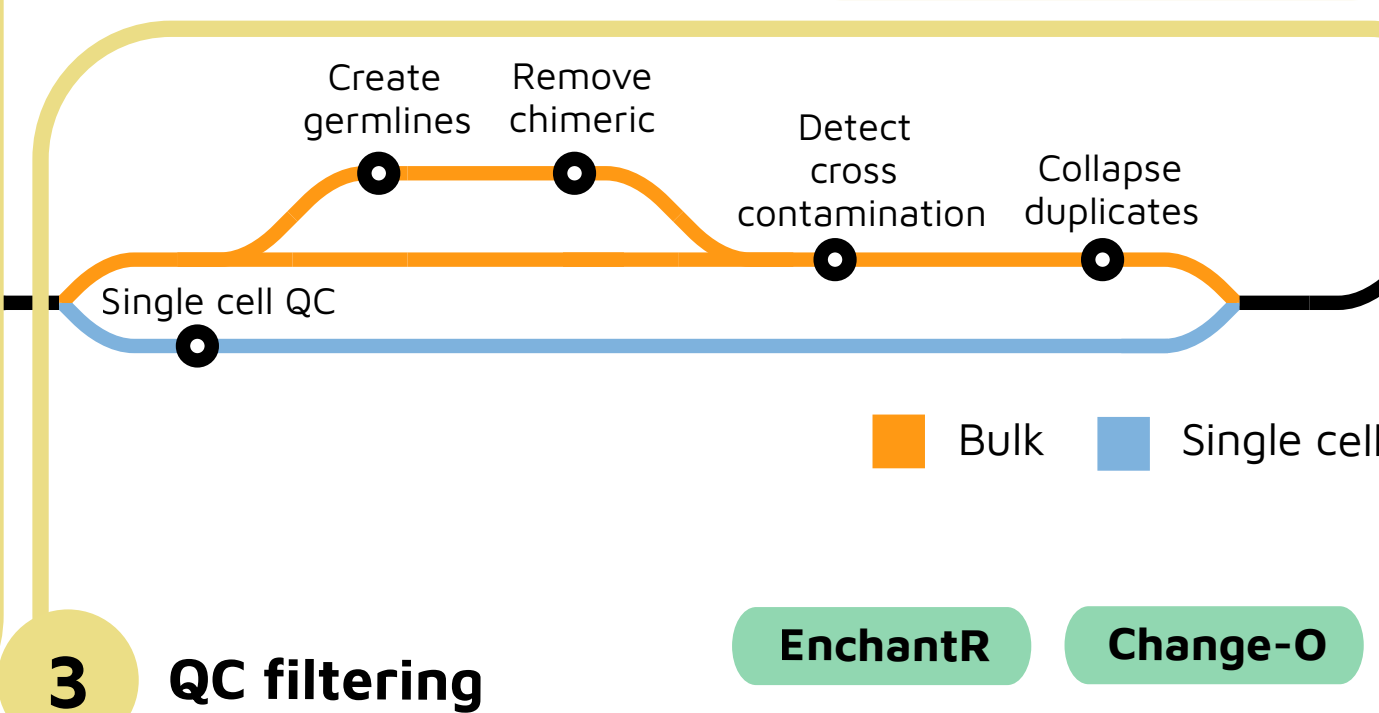
1 QC and sequence assembly (bulk)

Samplesheet tsv fastq
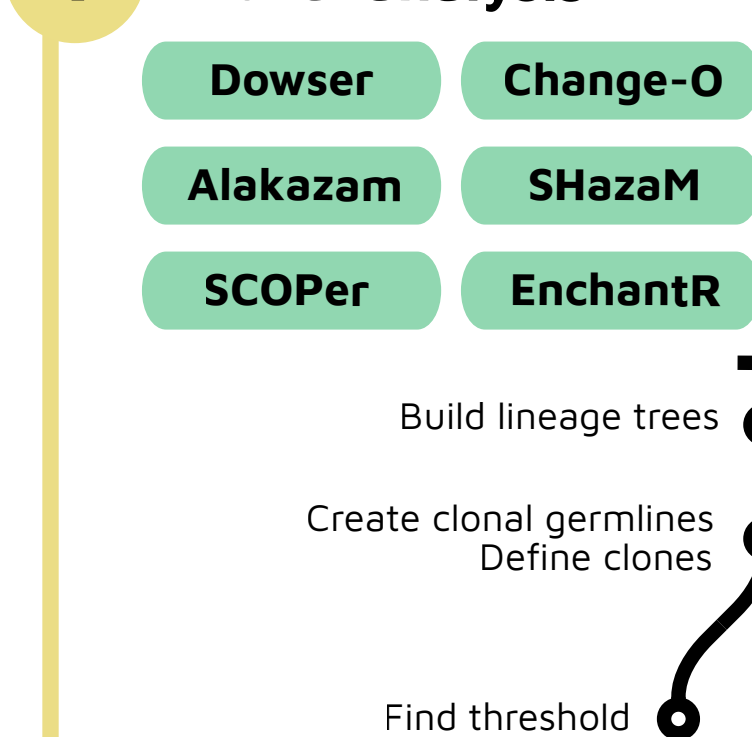
FastQC

Assemble pairs

Filter by quality

Mask primers cprimers vprimers

Pair sequences Cluster by similarity Annotate cluster ID

Build consensus

Pair sequences Assemble pairs

Collapse consensus count

FastQC

Annotate primers Annotate metadata Collapse duplicates Filter by duplicate count

fastQC pRESTO Sans-umi UMI-based Similarity clustering

2 V(D)J annotation and filtering

Samplesheet tsv fasta or AIRR

Convert to fasta IMGT

Assign genes

Fetch databases MakeDb

Filter quality

Productive

mod3 junction

Add metadata

EnchantR Change-O IgBLAST

3 QC filtering

Create germlines Remove chimeric Detect cross contamination Collapse duplicates

Single cell QC

Bulk Single cell

EnchantR Change-O

4 Clonal analysis

Dowser Change-O Alakazam SHazaM SCOPer EnchantR

Build lineage trees Create clonal germlines Define clones Find threshold

5 Repertoire analysis and reporting

Repertoire analysis QC reports html

Alakazam EnchantR MultiQC

nf-core/ airrflow
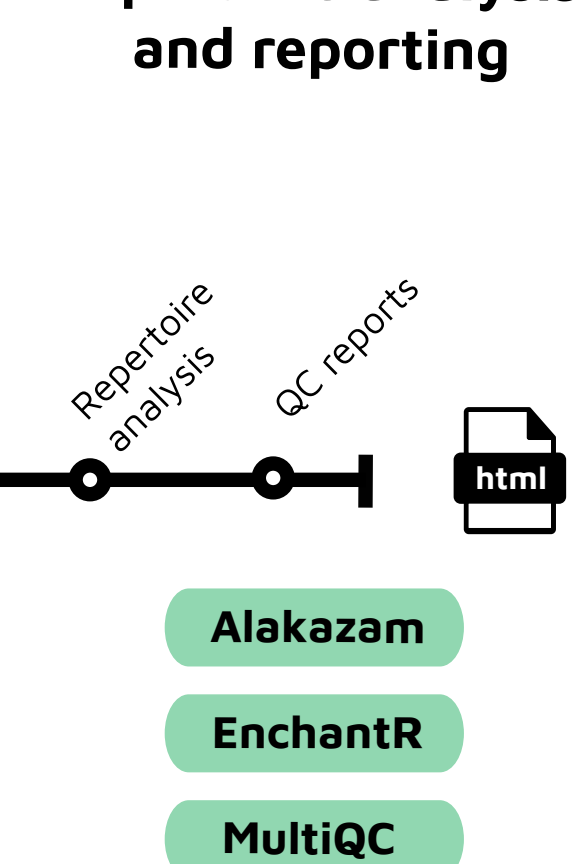
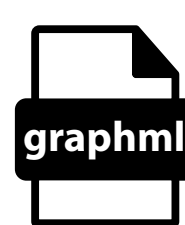## Outputs

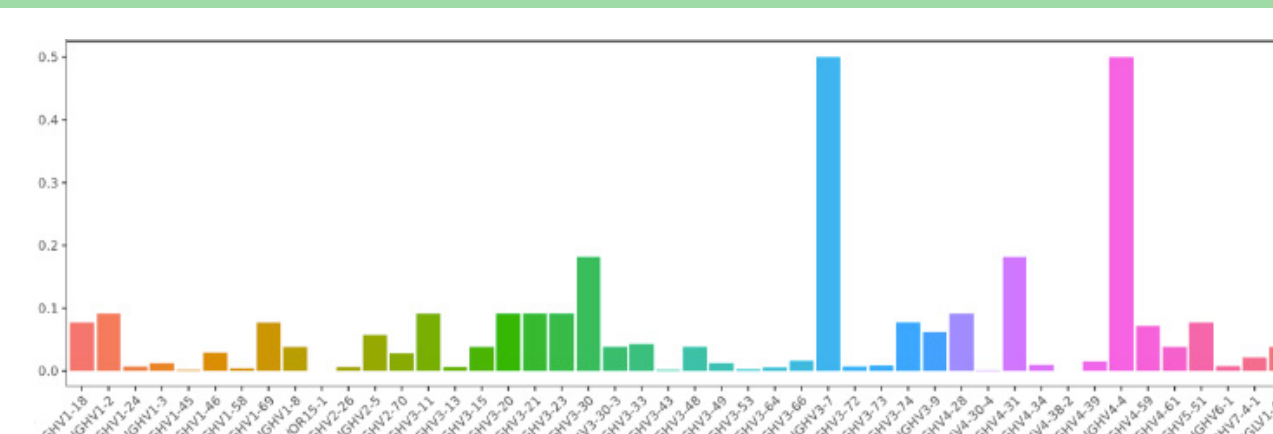AIRR format sequence rearrangements

Repertoire analysis reports

Lineage tree graphs
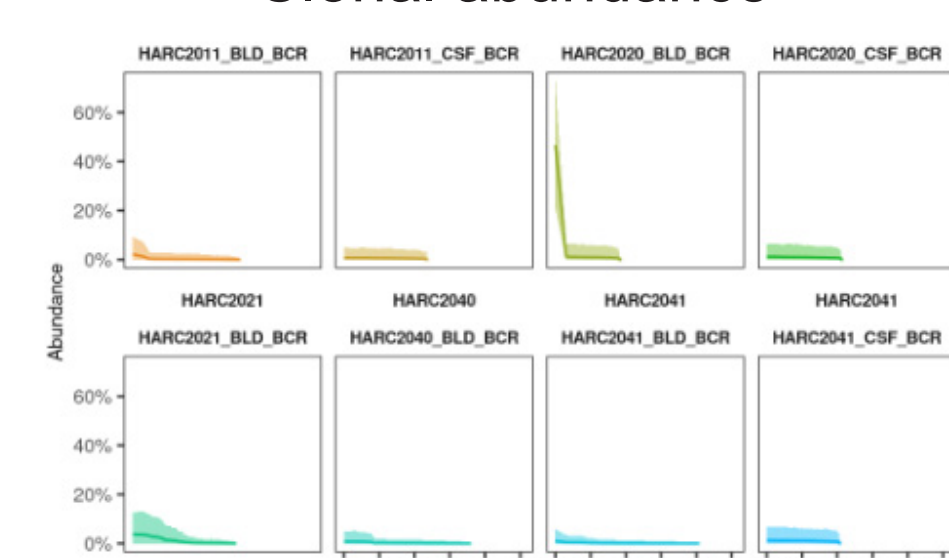
1 Input parameters
2 Read repertoires
  2.1 Sequences per locus
  2.2 Sequences per c_call
3 Clonal assignment
  3.1 Create germlines
4 Summary of clonal assignment
  4.1 Number of clones (heavy chain, li...
  4.2 Expanded clones within sample...
  4.3 Expanded clones within subject...
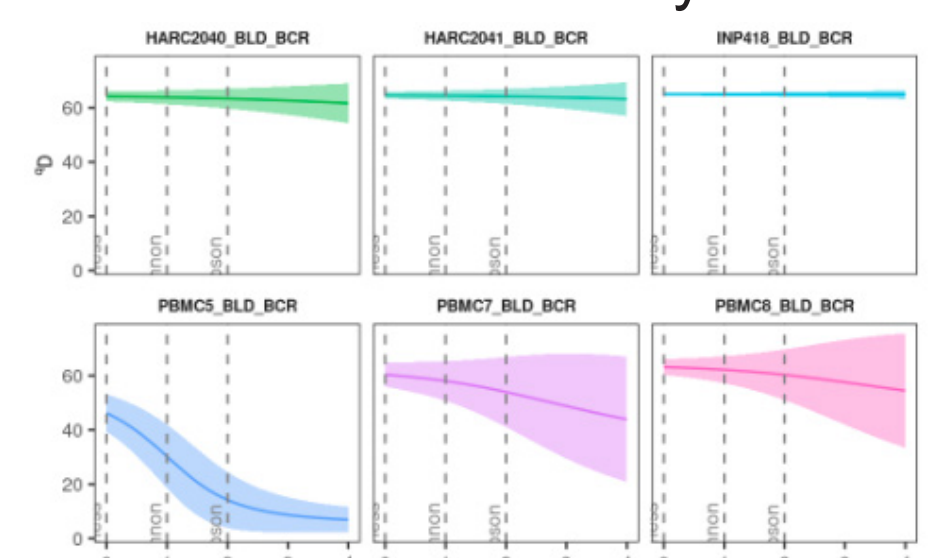 4.4 Clones spanning multiple tissues
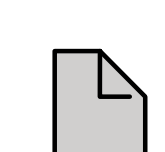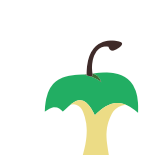5 Clone size distribution

V-gene usage

Clonal abundance

Clonal diversity

Get involved!

Milestones: today All processes in stable lifecycle status Incorporate scGEX in QC (Cell Ranger multi) Joint V(D)J and GEX analysis