Waipapa Taumata Rau
University of Auckland

ResBaz
RESEARCH BAZAAR

Reproducible Data Workflows with Snakemake

June 2026

Jean Love
Victor Gambarini

Housekeeping

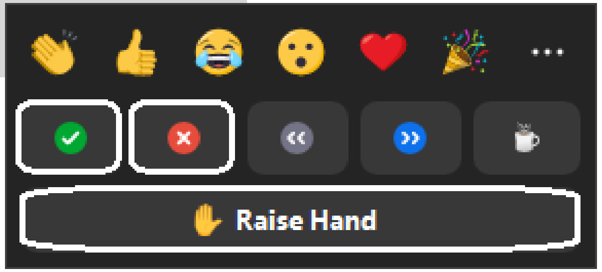
- Zoom: Stay **muted**, **camera on**
- **Change Zoom name** to your name (not UPI, etc.)
- **Participate** as much as possible

- **Introduce** yourself over Zoom chat (name, research area, institution, what you hope to get from this workshop)

- 2 hour code-along workshop with a 10 minute break just before the hour

- Use ✅ or ❌ To indicate you are following along

- Who has completed the Setup Instructions?



Questions and Help

- **Questions are good,** please put them in the chat if you don't understand something or are stuck.

- **Helper** will monitor the chat while the instructor is talking

- Instructors will have errors too! Wherever they come from errors are a good learning/debugging opportunity

- Some questions or errors will require we pause or handle them during the break

Snakemake

- A tool for creating reproducible and scalable data analysis **workflows**

- **Workflows** define a set of repeatable tasks

- **Workflows** are defined by **output files** created from **input files** linked together by **rules**

- The path of links between these rules to get to an output are **automatically determined** by Snakemake

- This allows you to run the same software pipeline with different inputs and setups

Snakemake

- Good For small scale pipelines you will be running often

- Based on Makefiles and Python

Other workflow languages:

- Common Workflow Language (CWL)
- **Nextflow - Excellent for larger scale pipelines!**
- Workflow Description Language (WDL)
- Guix Workflow Language (GWL)

Workshop overview

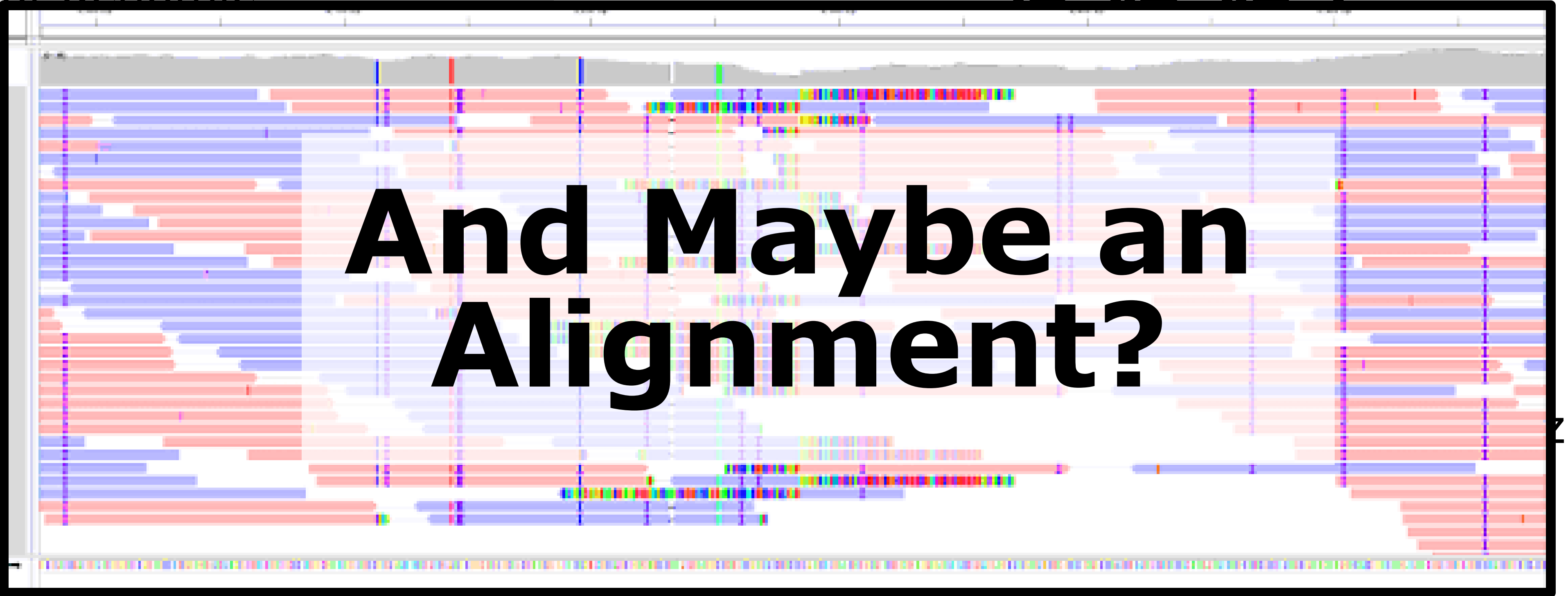
- An introduction to Snakemake using a basic Bioinformatics QC example
- Learn the structure of Snakemake rules
- Write run and assess a simple Snakemake workflow
- Extend the workflow with additional tools, logging and cores
- (if we have time) Define Snakemake configs

Thank you Leah Kemp, ESR and Genomics Aotearoa for developing and funding the **Workshop Content**

The Example

- Take a set of genomic sequencing
- Generate a of them (F/
- Perform qu each of the
- Aggregate for every fi summary (multiqc)

.fastq.gz



And Maybe an Alignment?

z

multiqc

Waipapa Taumata Rau University of Auckland

ResBaz RESEARCH BAZAAR

Start!

June 2026

Jean Love
Victor Gambarini

Waipapa
Taumata Rau
**University
of Auckland**

Break

June 2026

Jean Love
Victor Gambarini

Waipapa
Taumata Rau
University
of Auckland

ResBaz
RESEARCH BAZAAR

Thank You, Questions?

Thanks to Leah Kemp For developing this workshop content - original workshop For ResBaz 2020 funded by ESR and Genomics Aotearoa

June 2026

Jean Love
Victor Gambarini