

SorTR

Machine Learning for DNA Profile Analysis



Janet Stacey

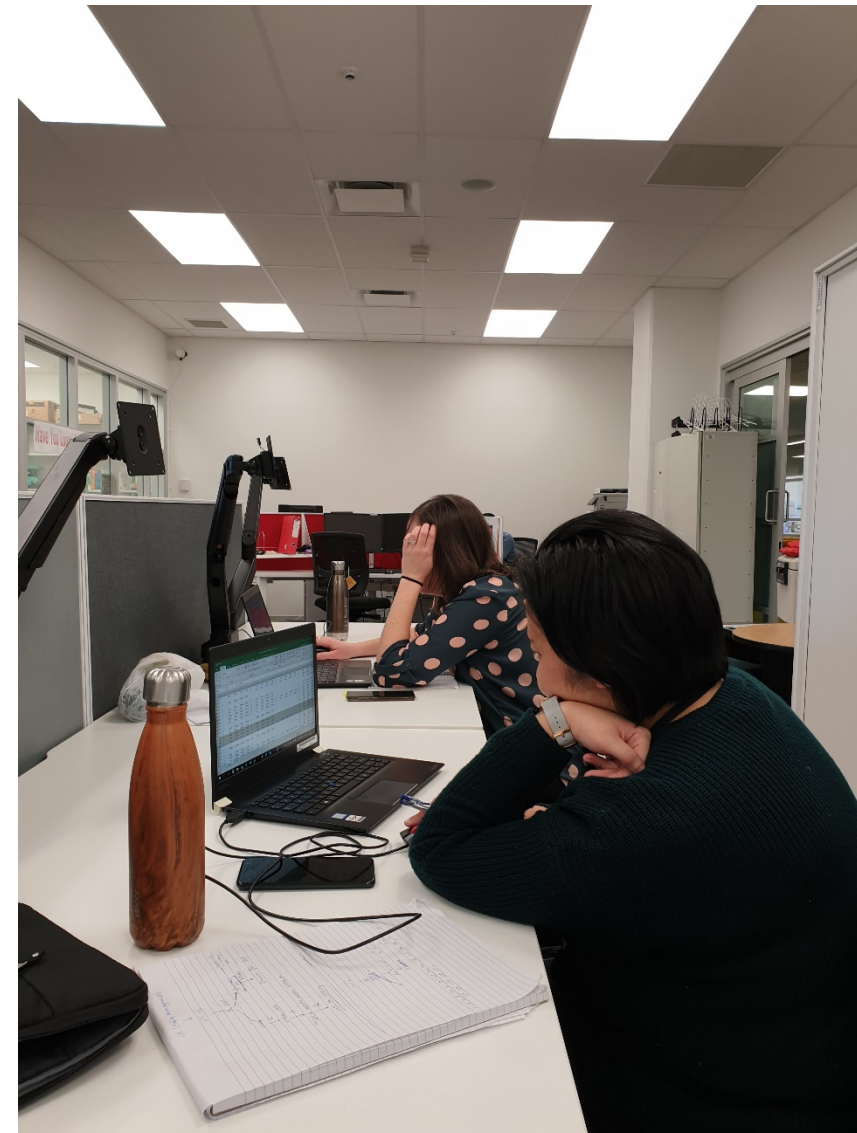
janet.stacey@esr.cri.nz

Anna Lemalu & Maria van der Salm

anna.lemalu@esr.cri.nz maria.vandersalm@esr.cri.nz

The Project

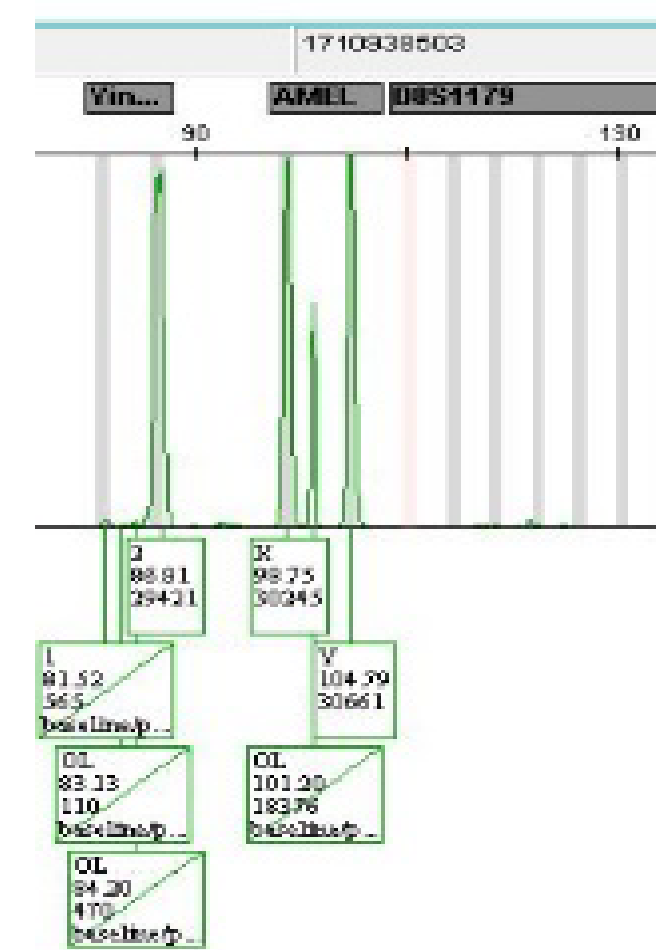
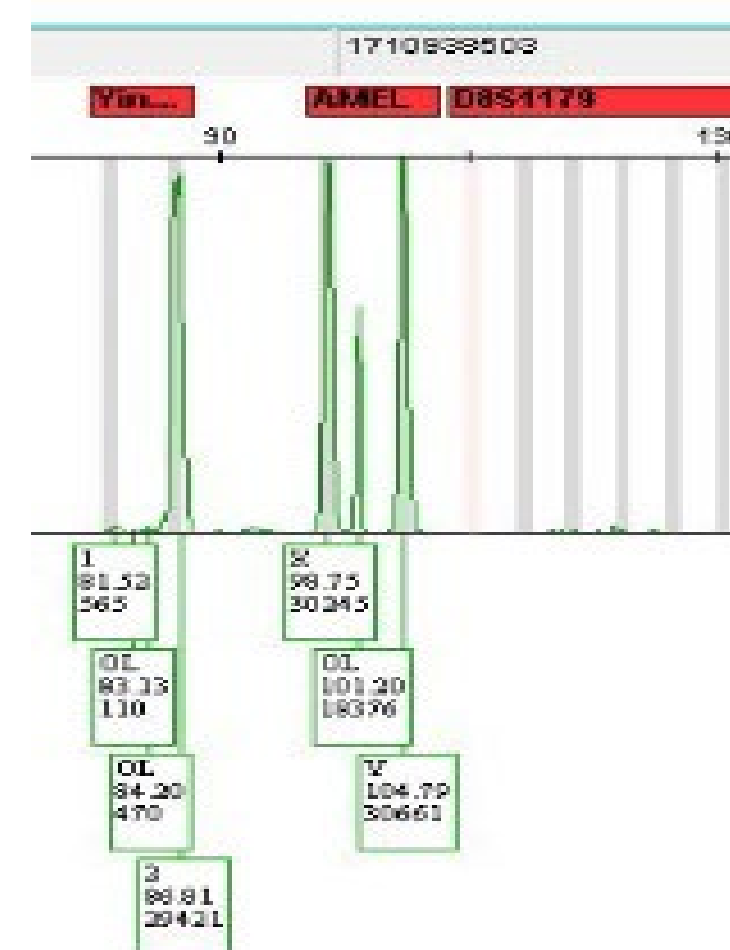
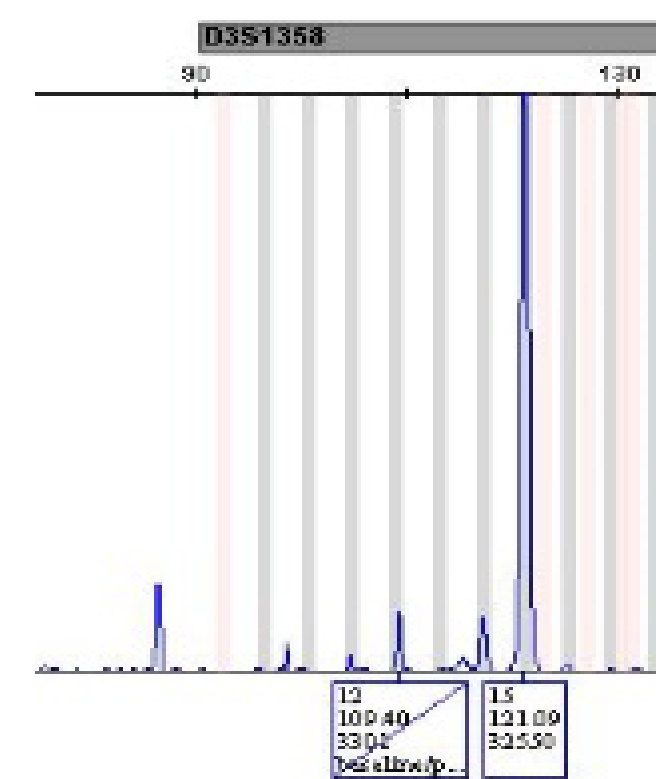
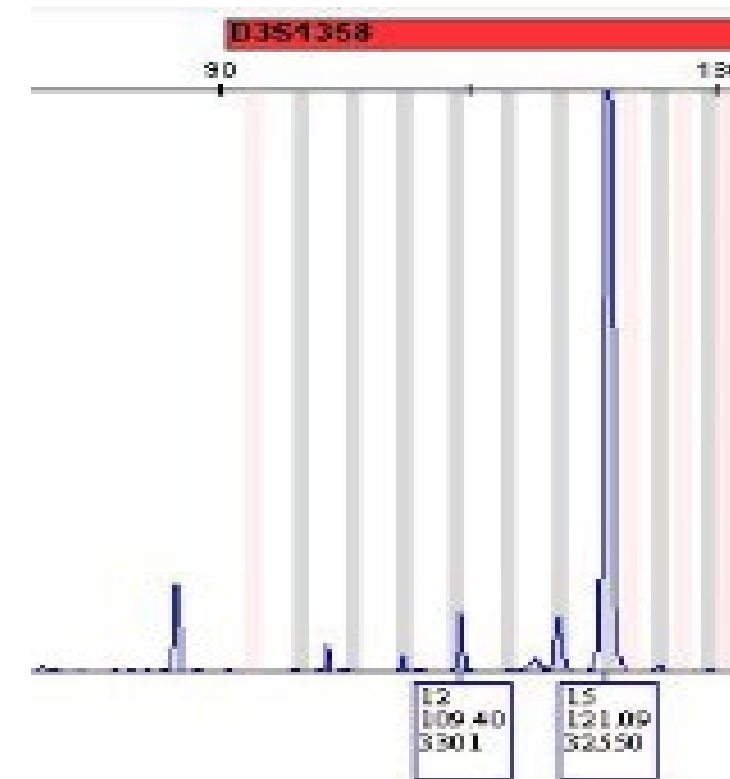
- Data Accelerator
- Cohort 1
- 15 weeks – 1 day a week



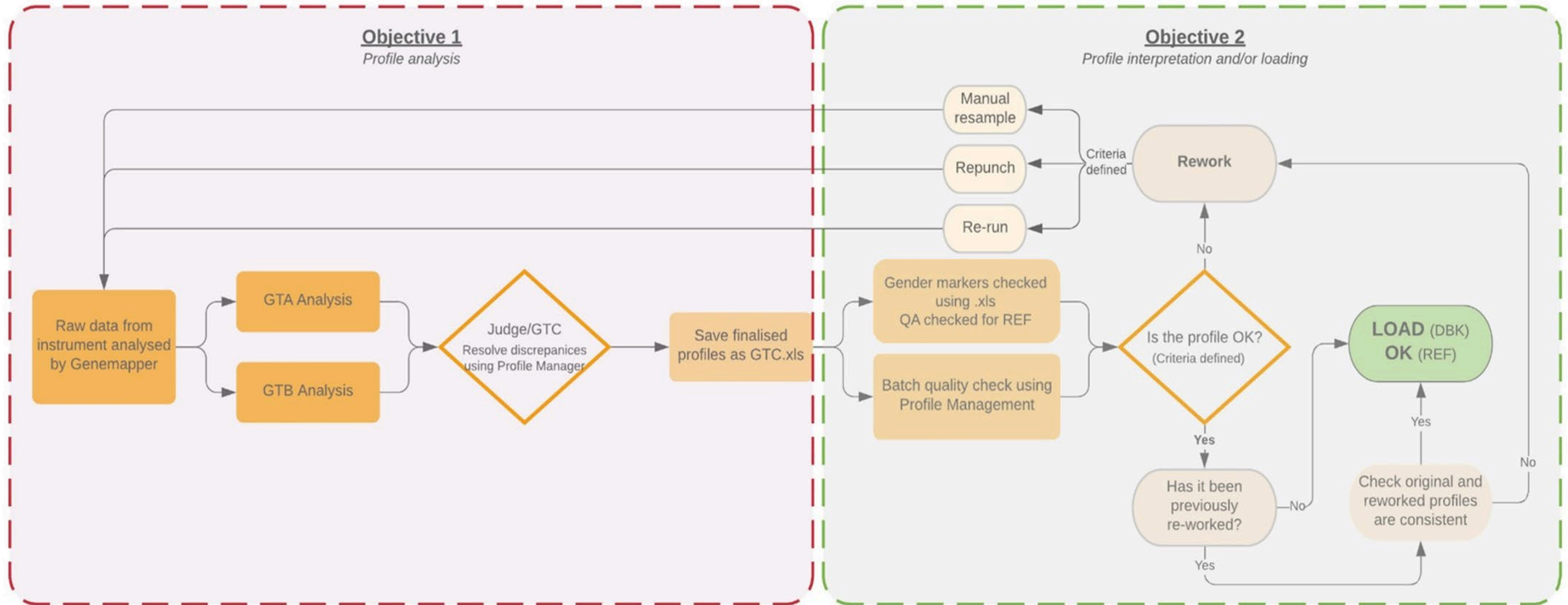
The Problem

- ~ 60 single source DNA profiles are analysed daily
- 3 different scientists and at least 3 hours of work per day

Marker	Allele 1	Allele 2
D3S1358	15	15
vWA	18	19
D16S539	9	11
CSF1PO	11	12
TPOX	8	11
Yindel	2	
AMEL	X	Y
D8S1179	13	13
D21S11	30	32.2
D18S51	16	25
DYS391	10	

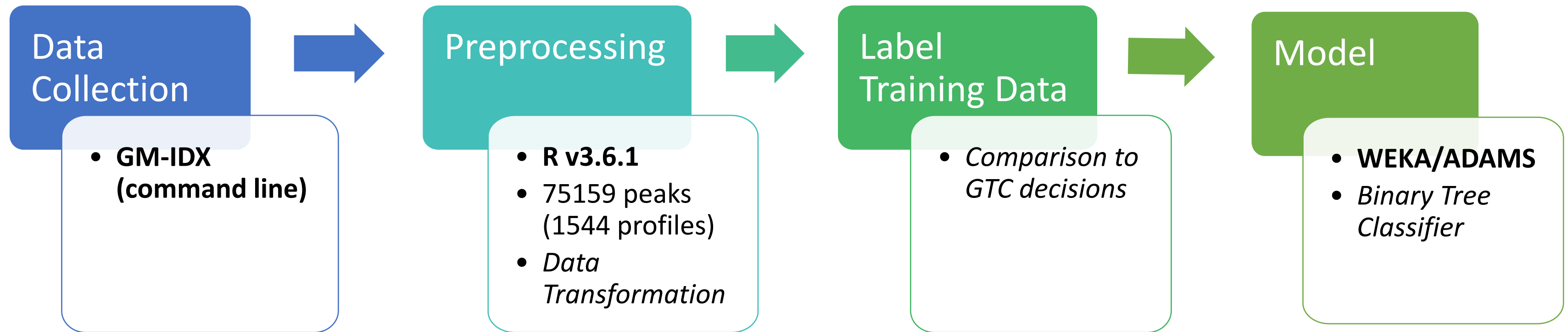


Current Workflow



Auto Allele Calling Classifier

- “*Is this peak an allele or an artefact?*”



3 → 2



1h → 8 minutes

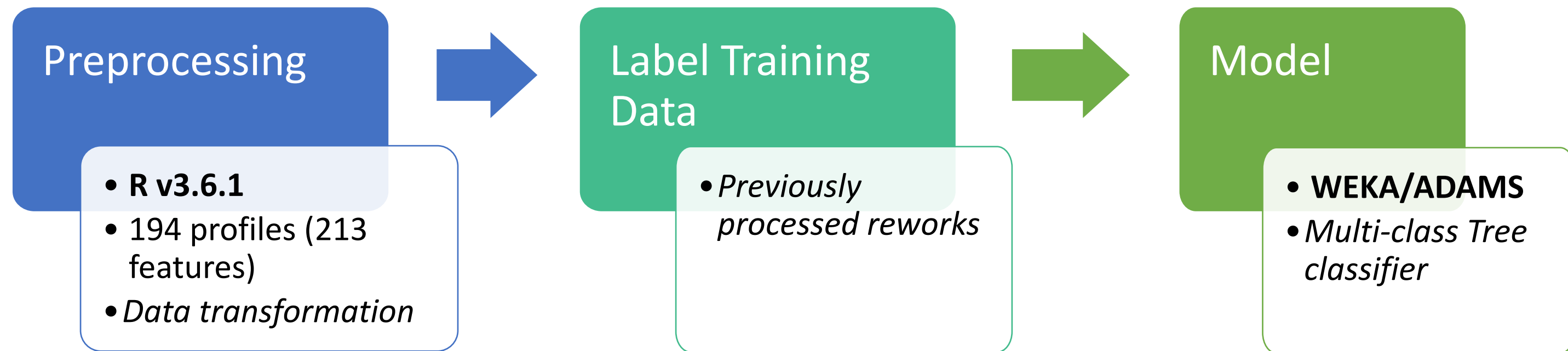


↓ Licence cost

Cross validated accuracy: 99.3%

Auto Rework Strategy Classifier

- *“Does this profile meet acceptance criteria?”*
- *If not - “Which rework method is best for this failed profile?”*



1h → ≤ 20 secs

Cross validated accuracy: 85%

Future work

- Further tweaks
- Validation
- Implementation
- Commercialisation

The SorTR Team



Anna

Senior Scientist (Forensic Biology)

- DNA Profiling Expert



Janet

Digital Science Engineer

**- Data Science &
Machine Learning Scientist**



Maria

Scientist (Forensic Biology)

- DNA Profiling Expert



Lucy

**Research Commercialisation
Advisor**