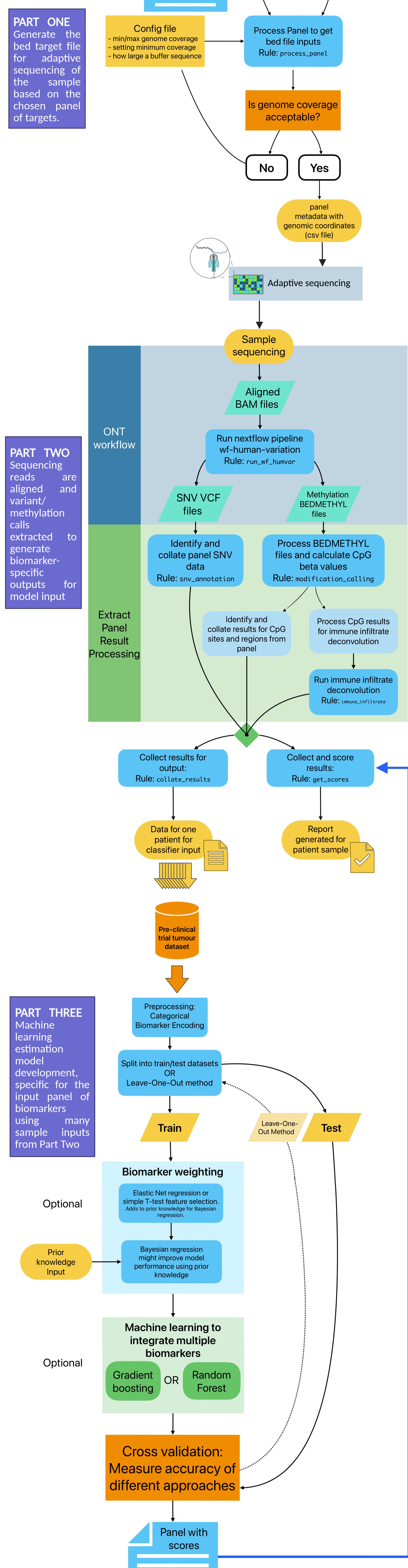


# Development of a multi-biomarker diagnostic pipeline

Lucy C. Picard<sup>1</sup>, Logan C. Walker<sup>2</sup>, Aaron J. Stevens<sup>1</sup>  
<sup>1</sup>University of Otago Wellington, <sup>2</sup>University of Otago Christchurch.

## PIPELINE



## INTRODUCTION

The broad goals of precision medicine are to improve quality of life, and enable an earlier commencement of the most beneficial treatment for patients. This pipeline addresses a key challenge in precision oncology: integrating heterogeneous biomarkers (mutations, methylation, and immune profiles) from nanopore sequencing data into a clinically interpretable framework.

The aim of this project was to develop a bioinformatic pipeline for nanopore-sequenced tumour samples that informs on treatment efficacy for a disease using multiple types of biomarkers. This proof-of-concept pipeline utilises the simultaneous detection of relevant DNA mutations and DNA methylation patterns enabled by nanopore technologies and additionally uses the produced methylation data to profile tumour-infiltrating immune cells. It was developed on the BCG treatment on non-muscle-invasive bladder cancer, with a simulated tumour sample, and is designed for use with any disease-treatment pair.

## INPUT PANEL

ID	Biomarker Name	Biomarker Type <sup>1</sup>	Panel or Area of Interest?	Chrom	Start pos	End pos	Length	Strand	Scoring Type	Result Options	Is Variant in Coding Region?	SNP ID	Is This Record the Whole Gene?	Illumina EPIC ID	DNA Methylation Region	Expression Ratio Components
004	NOS3	snv	Panel	chr7	150999023	150999023	1	+	genotypic	T  T T A T G A A A G G	Yes	rs1799983	No			
014	FLT1	mod <sup>2</sup>	Panel	chr13	2830346	28495128	194783	-	continuous	0.0-1.0	N/A	Yes				
017	IL6	exp_ratio <sup>3</sup>	Panel	chr7	22725884	22732002	6119	+	continuous	0.0-10.0	N/A	Yes				
018	IL10	exp_ratio	Panel	chr1	206768110	206774541	6432	-	continuous	0.0-10.0	N/A	Yes				
021	MMP10	expression	Panel	chr11	102770502	102780628	10127	-	continuous	0.0-1.0	N/A	Yes				
024	NOD2	area_mutations <sup>4</sup>	AOI <sup>7</sup>	chr16	50693606	50733075	39470	+	continuous	0.0-1.0	N/A	Yes				
034	IL18R1	microsatellite	AOI	chr2	101612009	101612009	1	+		GGGTGA	No	rs111894836	No			
401	LMR	immune_ratio <sup>5</sup>	Panel						continuous	0.0-10.0						
402	NLR	immune_ratio	Panel						continuous	0.0-10.0						
403	Monocyte_inf	immune_inf <sup>6</sup>	Panel						continuous	0.0-1.0						

**Table 1**  
Input panel to PART ONE of the pipeline, sample of possible entries. Along with the genomic coordinates (to hg38), optional secondary data is available where applicable and increases efficiency (for example, SNP ID, Illumina EPIC ID).

<sup>1</sup> Biomarker Type: can be one of snv, sv, mod, area\_mutations, expression, exp\_ratio, immune\_ratio, immune\_inf or microsatellite. <sup>2</sup> mod: DNA modification.

<sup>3</sup> area\_mutations: any mutations in the specified area influence BCG efficacy. <sup>4</sup> exp\_ratio: expression ratio between two gene entries. <sup>5</sup> immune\_ratio: ratio of specific immune cell types from the infiltrate estimation. <sup>6</sup> immune\_inf: infiltrate of the specified immune cell type. <sup>7</sup> AOI: Area of Interest.

## RESULTS FROM A TEST SAMPLE

**Sample sequencing:** A simulated sample was constructed: 25% bladder cancer cell line and 75% leukocytes<sup>4</sup>. DNA was extracted using the Qiagen Blood and Tissue Kit, and sequenced using the Rapid Barcoding Kit (RBK114.24) on a P2 Solo. Adaptive sequencing performed using a bed file (PART ONE), from a minimal panel of biomarkers relevant to BCG therapy. Sequences were basecalled using the EPI2ME Labs workflow wf-basecalling (v1.4.1), then used as input to PART TWO of the pipeline.

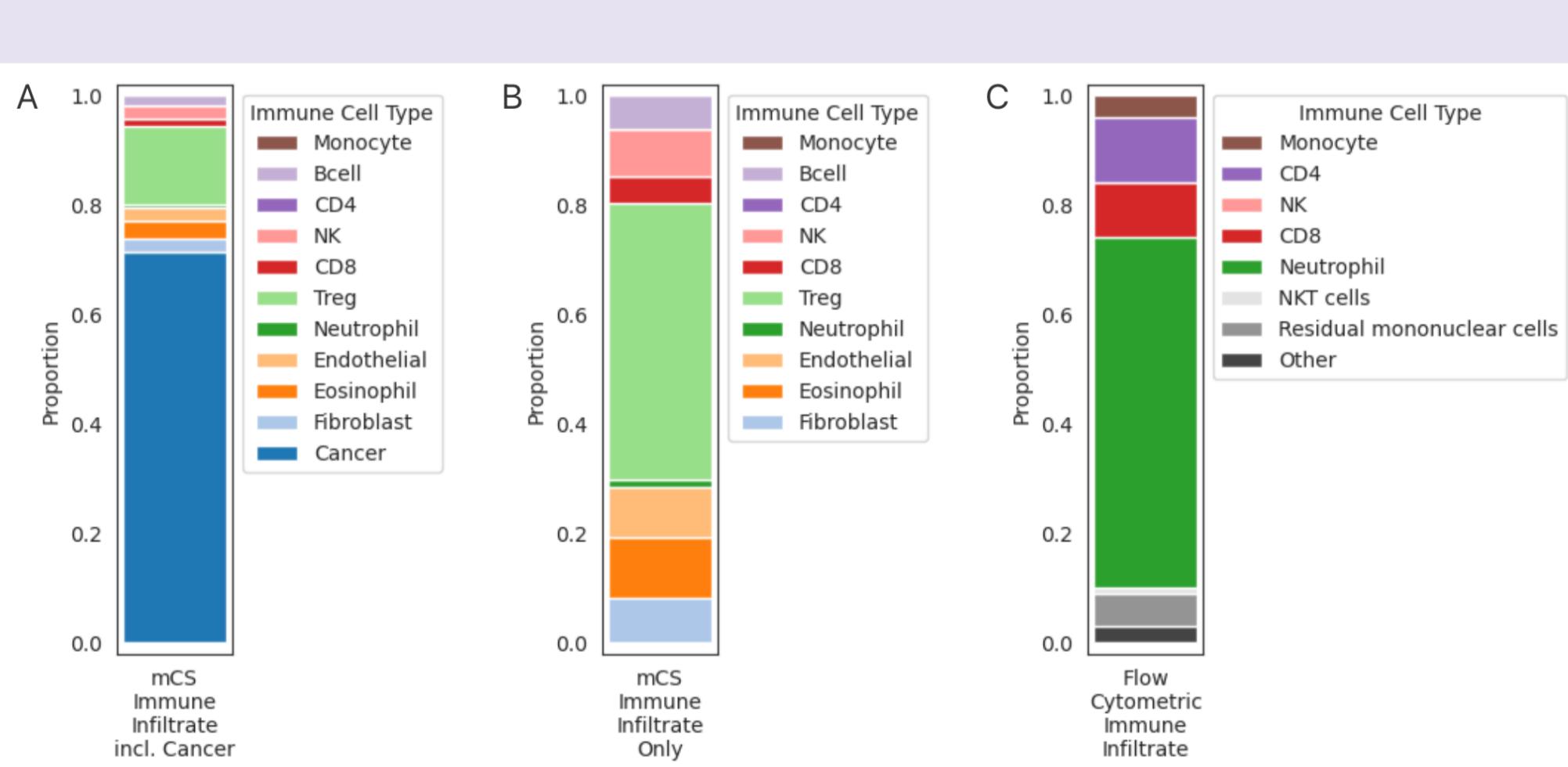
**SNV and methylation findings:** SNV and methylation results were obtained from EPI2ME Labs wf-human-variation pipeline and was straightforward to integrate into the pipeline, producing sufficient "panel output" results (Table 2).

**Immune deconvolution limitations:** Immune infiltrate results did not reflect the 25/75 proportion of cancer cells/leukocytes, nor the immune cell composition expected from a healthy donor (Table 2, Figure 1). Monocytes were not detected at all, prohibiting the Lymphocyte to Monocyte ratio (LMR) calculation, important to the prediction of many immunotherapeutic treatments. Similarly, neutrophils and CD4+ T-cells were absent, while T-reg cells were estimated to be the majority (14%) of the leukocytes present. The failure of immune infiltrate prediction could be due to a number of factors, including an incomplete or inaccurate reference, or an unsuitable sample composition for mCS. These could be addressed by using a sequencing-based deconvolution reference, as methods between array-based and sequencing-based algorithms differ<sup>5</sup>, and by using a real tumour sample.

ID	Biomarker name	Scoring Type	Biomarker Type	Result Options	Result	Score
004	NOS3	genotypic	snv	T  T T A T G A A A G G	G G	0.16
014	FLT1	continuous	mod - promoter region	0.0-1.0	0.08	1.55
401	LMR	continuous	immune_ratio	0.0-10.0	NA	
402	NLR	continuous	immune_ratio	0.0-10.0	0.02	1.66
403	Monocyte_inf	continuous	immune_inf	0.0-1.0	NA	
405	Bcell_inf	continuous	immune_inf	0.0-1.0	0.02	1.72
406	CD4_inf	continuous	immune_inf	0.0-1.0	0	0.33
407	NK_inf	continuous	immune_inf	0.0-1.0	0.02	
408	CD8_inf	continuous	immune_inf	0.0-1.0	0.01	0.07
409	Treg_inf	continuous	immune_inf	0.0-1.0	0.14	1.52
410	Neutrophil_inf	continuous	immune_inf	0.0-1.0	0	
411	Endothelial_inf	continuous	immune_inf	0.0-1.0	0.03	
412	Eosinophil_inf	continuous	immune_inf	0.0-1.0	0.03	
413	Fibroblast_inf	continuous	immune_inf	0.0-1.0	0.02	
414	Cancer_inf	continuous	immune_inf	0.0-1.0	0.71	
Total (raw)						33.5
Total (normalised to 0.0 - 100.0)						47.05

**Table 2**  
Final result from simulated sample processing by the full pipeline with biomarker "scoring" added. Hazard ratio per biomarker is determined from PART THREE model results, and the "score" is computationally determined from the z-scores of each biomarker.

**Figure 1: Immune infiltrate of composite sample.**  
(A) Total immune proportion, including cancer cells detected by MethylCIBERSORT and CIBERSORT (mCS). (B) Immune infiltrate proportion from (A) excluding cancer cells. (C) Equivalent immune proportions from the leukocyte sample detected by flow cytometry. Grey shades indicate groups that are not included in those present in the MethylCIBERSORT reference. (A) and (B) are output during Part Two and Part Four of the pipeline.



## FOOTNOTES

- Chakravarthy, A. et al. Pan-cancer deconvolution of tumour composition using DNA methylation. *Nat Commun* 9, 3220 (2018).
- Newman, A. M. et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol* 37, 773–782 (2019).
- Plaisance, S., VIB Belgium, Creative Commons Attribution-ShareAlike 3.0 Unported License
- Explicit permission was obtained from the healthy leukocyte donor for this purpose.
- De Ridder, K., Che, H., Leroy, K. & Thienpont, B. Benchmarking of methods for DNA methylome deconvolution. *Nat Commun* 15, 4134 (2024).

## ACKNOWLEDGEMENTS

This work was funded by the Health Research Committee of NZ and the Otago Doctoral Scholarship. Travel funding to present this work was awarded from the Cancer Society of New Zealand and the Maurice and Phyllis Paykel Trust. Advice for adaptive sequencing received from Leticia Castro and Jane Clapham, ESR, NZ.

## DISCLAIMER

Oxford Nanopore Technologies products are RUO. Products labelled/branded as Oxford Nanopore Diagnostics may be RUO or may be regulated as in-vitro diagnostic devices in some jurisdictions, please check individual product labelling. Oxford Nanopore Technologies present the materials of third parties within the user community acknowledging that those materials belong to and are the true statements and representations of those third parties. Oxford Nanopore Technologies make no endorsements of the materials shown and the statements made. There are no conflicts of interest.

## NEXT STEPS

Further additions to the software, such as characterising gene fusions, are easy to implement given the modular design of Snakemake. Immediate next steps are to optimise the wet-lab processing and sequencing of real tumour samples, and ensure an accurate immune infiltrate result can be obtained.



Project on GitHub