**Project title:** Detection of aberrantly expressed genes and pathways in tumor sequencing for precision oncology

**Public Research Use Statement**

We seek to detect aberrantly expressed genes and altered pathways in transcriptome data derived from cancer patients in order to describe their molecular profile, identify actionable oncogenes and predict treatment response. Our standard transcriptome sequencing pipeline, however, does not include sequencing matched normal samples. Thus, we plan to use The Genotype-Tissue Expression (GTEx)1 normal sample data derived from the same tissue as controls to complement the genome-directed discovery with transcriptome profiling in the context of routine precision medicine sequencing performed in University of Melbourne Centre for Cancer Research (UMCCR).

**Technical Research Use Statement**

Our objective is to improve the molecular detection and diagnosis of various cancer types, prioritise therapeutic selection and enabling programs in personalised cancer care. To achieve this, we generate different levels of high-throughput biological data, such as whole-genome and transcriptome, in the context of routine precision medicine sequencing. Currently, our patient reporting system is based on genome-directed discovery, which we plan to complement with transcriptome profiling to aid identification of aberrantly expressed genes, deregulated pathways, oncogenic fusion genes, as well as stratification of patients into relevant molecular groups.

The transcriptomics data are analysed using bcbio RNA-seq2 pipeline (https://bcbio-nextgen.readthedocs.io/en/latest/contents/pipelines.html#rna-seq) with STAR3 used for initial alignment, Salmon4/Kallisto5 for quantification and Pizzly6/Oncofuse7 for detecting gene fusions. To identify aberrantly expressed genes and deregulated pathways, however, we need control samples derived from the same tissue. Hence, we plan to process GTEx1 raw reads, derived from normal tissue corresponding to the tumor type, with bcbio RNA-seq pipeline to obtain uniformly processed data that can be used as control cohort. This integrative approach will facilitate determining comprehensive molecular and genetic profile of each patient and ultimately will tailor medical treatment to their individual characteristics.

**References**

1. GTEx Consortium, Genetic effects on gene expression across human tissues. *Nature* 2017; 550(7675):204-213
2. Steinbaugh MJ, Pantano L, Kirchner RD *et al*. bcbioRNASeq: R package for bcbio RNA-seq analysis. *F1000Research* 2018; 6:1976
3. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013; 29(1):15-21
4. Patro R, Duggal G, Love MI, Irizarry RA and Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods* 2017; 14(4):417-419
5. Bray NL, Pimentel H, Melsted P and Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology* 2016; 34(5):525-527
6. Melsted P, Hateley S, Joseph IC, Pimentel H, Bray NL, Pachter L. Fusion detection and quantification by pseudoalignment. *bioRxiv* 2017; doi: <https://doi.org/10.1101/166322>
7. Shugay M, Ortiz de Mendíbil I, Vizmanos JL and Novo FJ. Oncofuse: a computational framework for the prediction of the oncogenic potential of gene fusions. *Bioinformatics* 2013; 29(20)2539-2546