PM-OICR TGL Manual for Grant Applications, Proposals and Clinical Research Protocols

Abstract

This document is a supplement to our website and lab protocols. Lab and informatic protocols are summarized for easy insertion into grant applications, proposals and clinical protocols for ethics review committees (CAPCR, iREB). Full lab and informatic protocols and workflow diagrams are available at https://labs.oicr.on.ca/translational-genomics-laboratory/

Contents

1. PM-OICR TGL Mandate:

The Princess Margaret Cancer Centre-Ontario Institute for Cancer Research Translational Genomics Lab (PM-OICR TGL) is a joint collaborative workspace located in the MaRS complex at University Avenue and College Street in downtown Toronto. TGL enhances rapid access to genomics technologies, and facilitates genomic interpretation and reporting in clinical oncology. TGL provides research support to the Princess Margaret Cancer Centre and affiliated cancer research institutions as part of OICR's Adaptive Oncology Program.

TGL is formulated to accelerate clinical oncology research through partnership and drive key initiatives:

- Facilitate access to genome-wide multi-omic assays from pathology specimens (Formalin Fixed Paraffin Embedded, FFPE);
- Identify molecular patterns associated with patient outcome and clinical variables from standard of care and second line therapies;
- Share and continuously improve integrated analysis methods to facilitate robust tumor phenotypic classification;
- Facilitate actionable mutation detection through an emphasis on multiplatform molecular diagnostics, including methylation, transcriptome and exome sequencing;
- Integrate variant reporting into work streams through cBioPortal and Heliotrope, allowing rapid, automated dissemination of results to the oncologic community and research databases, including ICGCmed

2. Pathology Lab Requests (Tumor Tissue):

PM-OICR TGL recommends the following parameters for pathology tissue lab requests: 10 micron sections, minimum n=12, 1 H&E at the top of the sectioning stack to establish tumor cellularity, 10 unstained slides cut onto uncharged slides (if study includes immunohistochemistry assays, charged slides are required), and 1 H&E at the end of the sectioning stack to verify tumor cellularity. If there are discrete areas of tumor, please request the pathologist to circle regions on the two requested H&E slides and score the % tumor cells, % necrosis for the circled region. If the entire tissue section is tumor, the entire tissue may be circled and scored. Ideally at total surface area of >150 mm^2 (15mm^2 tumor surface area X 10 slides) will yield sufficient DNA and RNA for exome and RNA sequencing libraries; macro dissection of multiple slides may be required. A maximum of 600mm^2 of tumor tissue (100mm^2 tumor surface area X 6 slides) may be extracted over 1 purification column set. Tumors with a total surface area of <150mm^2 should be prioritized for DNA or RNA extraction only; separate

protocols apply. This information may be detailed in PM-OICR TGL submission sheets. De-identified study codes must be used on all documentation.

3. Lab Protocols

3.1 DNA Isolation from Buffy Coat

DNA is extracted from 150-250 ul of buffy coat using the Qiagen Gentra Purgene Blood Kit according to manufacturer's directions.

3.2 DNA/RNA Co-isolation from FFPE Slides Protocol

DNA and RNA is co-isolated from 150mm^2-600mm^2 of macro dissected tumor surface area from 10 micron sections. Macro dissected material is deparaffinized using CitriSolv reagent, proteinase K digested, and DNA pellets and RNA supernatant purified over Qiagen AllPrep DNA and RNA kit columns. Isolated DNA is RNase treated, and RNA is DNase treated. Isolated material is suitable for all sequencing protocols. Full protocol details and modifications for FFPE RNA and DNA co-isolation are available on our website.

3.3 RNA Sequencing Lab Protocol:

RNA libraries are synthesized from 200 ng of Total RNA using the Illumina TruSeq Stranded Total RNA Sample Prep kit. Total RNA is depleted of ribosomal RNA, first and second strand cDNA is synthesized, A-tailed, adapter ligated, and PCR amplificated. Full protocol details and modifications for FFPE RNA are available on our website. Tumor RNA is sequenced on the Illumina NextSeq550 platform, V2 Chemistry and reagents, to read depth of 80 million clusters, 160 million paired end reads, 75bp X 75bp.

3.4 Exome Sequencing Lab Protocol

Exome libraries are prepared from 100 ng of DNA from fresh frozen or FFPE tumor material, and/or normal buffy coat DNA. Pre-capture libraries are synthesized using a modified protocol based on the KAPA Hyper Prep Kit, prior to capture using a modified Agilent XT V6 + COSMIC exome workflow. Briefly, DNA is sheared, prior to end repair, A-tailing, adapter ligation, and PCR amplification. Exome probes are hybridized to DNA libraries for 16 hours and washed, prior to on-bead amplification and cleanup. Full protocol details are available on our website. Matched normal DNA is sequenced to a depth of 50X coverage, Tumor DNA, 250X coverage on the Illumina HiSeq2500 platform, V4 Chemistry and reagents, with read lengths of 125bpX 125bp.

3.5 Methylation EPIC Exome Sequencing Lab Protocol:

TruSeq Methyl Capture EPIC exomes are prepared from 500ng of tumor DNA according to manufacturer's directions. DNA is fragmented, end repaired, A-tailed, and adaptors ligated. Methylation EPIC probes are hybridized overnight, prior to bisulfite conversion, PCR amplification and cleanup. Libraries are sequenced on the Illumina HiSeq2500 platform, V4 Chemistry and reagents, with read lengths of 125bpX125bp.

3.6 Methylation Array Protocol:

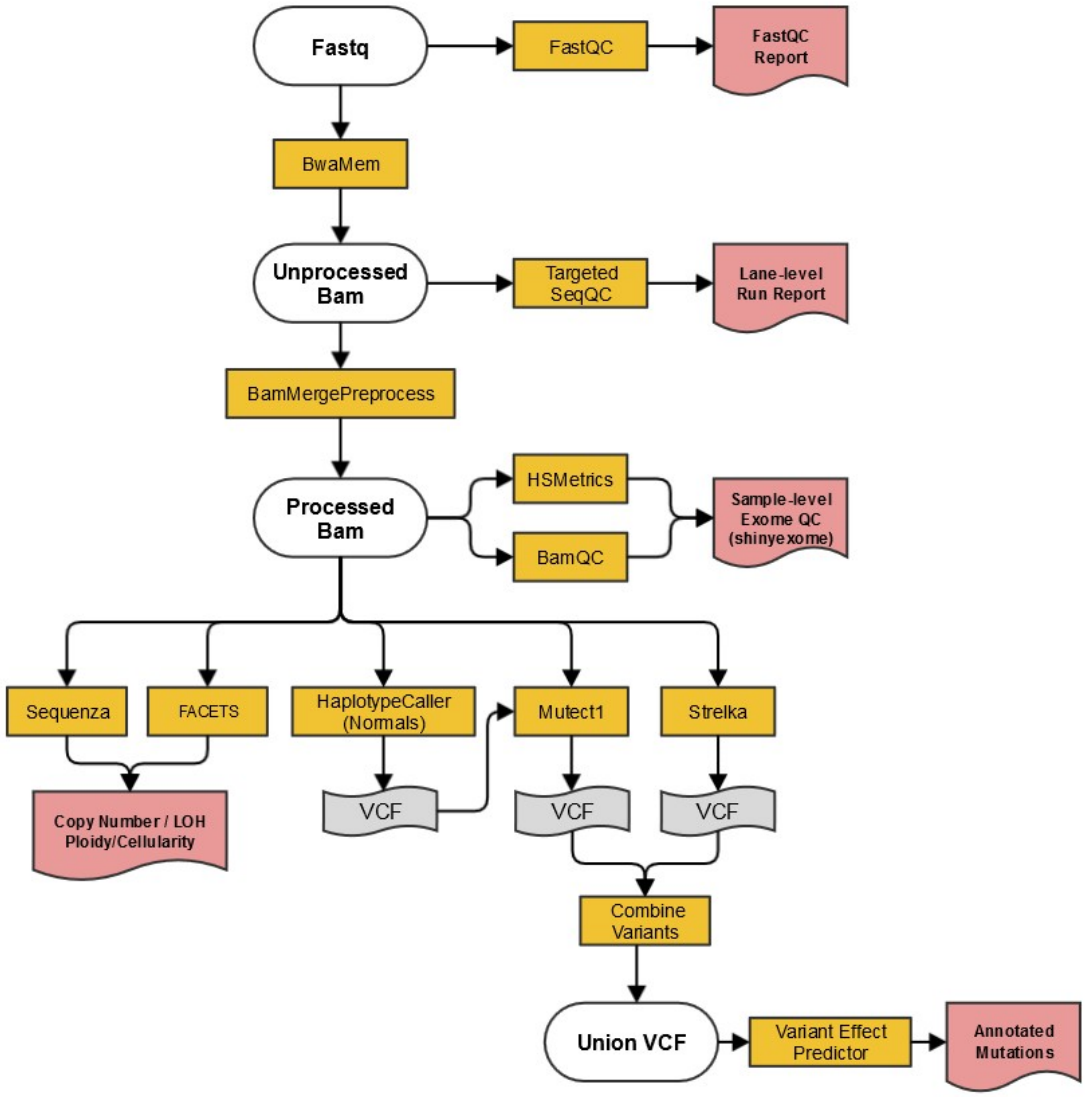
Illumina Infinium Methylation EPIC BeadChip arrays are processed according to Manufacturer's instructions. Briefly, 250 ng of DNA (from fresh frozen or FFPE) is bisulfite converted using EZ DNA Methylation kit (Zymo research), FFPE DNA is repaired (Infinium HD FFPE Restore, if applicable), followed by whole genome isothermal amplification, enzymatic fragmentation, DNA precipitation and resuspension. Prepared samples are loaded on beadchips and hybridized for 16-20 hours at 48°C. Hybridized samples are washed prior to single-base extension with labelled fluorophores, followed by washing and scanning on the Illumina ISCAN system.

4. Informatics

4.1 Exome Informatics:

Sequence reads will be aligned against hg19 human reference using BwaMem v 0.7.12 to generate raw sequence alignments in BAM format. Preprocessing which includes PCR-duplicate marking, indel re-alignment and base quality recalibration will be performed using Picard v1.72 and GATK v3.6.0 [1]. When available, preprocessing will be performed in matched tumor/normal pairs to improve indel re-alignment. Mutect1 v1.1.7 [2], Samtools v0.1.19 [3], and Strelka v1.0.13 [4] will be run to create raw variant call files (VCFs).

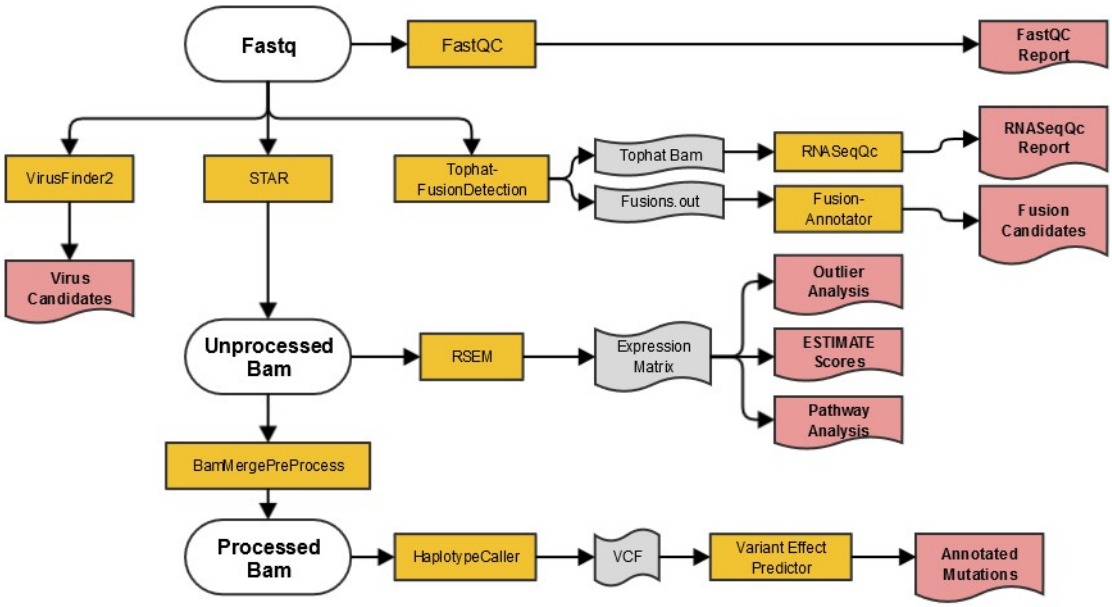
Raw VCF files will be annotated with Variant Effect Predictor v83 [5] against multiple variant annotation databases. Variants will be annotated with GnomAD allele frequencies [6] – a database of variant frequencies in a healthy population – in order to remove common variants. Variants will also be annotated against known hotspots in cancer (CancerHotspots.org) both at the variant level, and gene level [6, 7]. Analysis will include actionable /oncogenic driver analysis using the Precision Oncology Knowledge Base (oncoKb) and pathogenic database ClinVar [8, 9] . Through this analysis, actionable variants will be classified according to standard therapeutic intervention, investigational therapeutic implication, hypothetical therapeutic intervention, and standard therapeutic implications (resistance) including oncodriver annotations of inconclusive, likely neutral, likely oncogenic, oncogenic and unknown. Additional analysis may be applied to detect allele specific copy number profiles, ploidy/cellularity (Sequenza)[10] and estimates of immune infiltrates with either Immunomap (T-cell receptor sequence diversity)[11] and Cibersort packages [12]. Clinical protocols that allow for data sharing as defined in patient consent forms, or as approved by ethical review boards may use a secured instance of cBioportal [13, 14] for visualization of genomic data cohorts.



4.2 RNA Informatics:

PM-OICR TGL will process Total RNA through STAR aligner v2.5.1b [15] followed by RNA abundance quantification using RSEM v1.3.0 [16] generating an expression matrix used for expression outlier analysis, ESTIMATE [17] immune infiltrate analysis, and ssGSEA [18] pathway analysis. For calling variants, BAM files will be preprocessed similar to exome methods, except an additional trimming of soft-clipped reads is performed prior to indel re-alignment and base recalibration. Haplotype Caller will generate variant call files (VCF), prior to variant effect predictor analysis of annotated mutations [1]. Tophat Fusion Detection v2.0.10 and STAR-fusion [19] will be run for detection of fusion candidates.

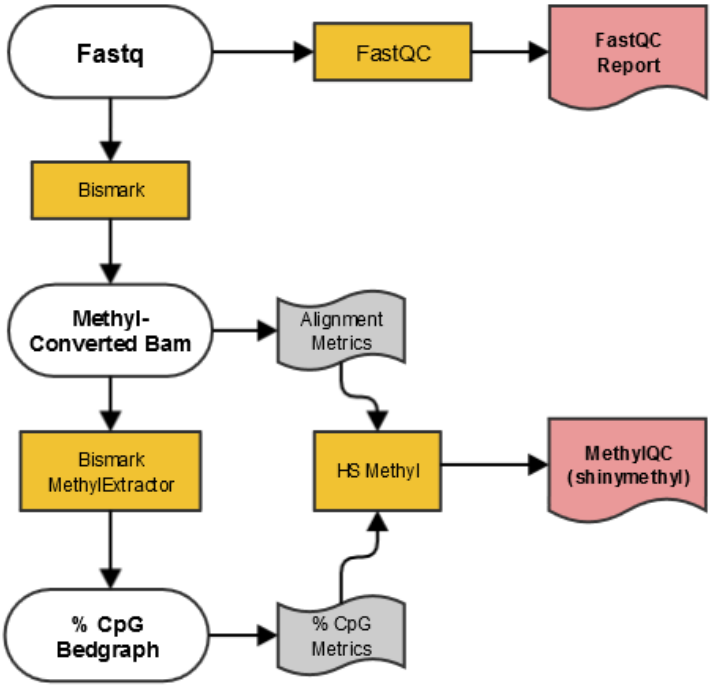
Clinical protocols that allow for data sharing as defined in patient consent forms, or as approved by ethical review boards may use a secured instance of cBioportal [13, 14] for visualization of genomic data cohorts.



4.3 Methylation Sequencing Informatics:

Sequence reads from Illumina TruSeq Methylation EPIC exome libraries will be aligned using Bismark software v0.15.0 [20]. Methyl-converted BAM files will be processed using Bismark Methyl Extractor generating bed graph files of % CpG methylation for each site. Library quality metrics including mapping efficiency, % reads on target, and distribution of % CpG methylation across samples will be calculated. Clinical protocols that allow for data sharing as defined in patient consent forms, or as approved by ethical review boards may use a secured instance of cBioportal for visualization of genomic data

cohorts. Clinical protocols that allow for data sharing as defined in patient consent forms, or as approved by ethical review boards may use a secured instance of cBioportal [13, 14] for visualization of genomic data cohorts.



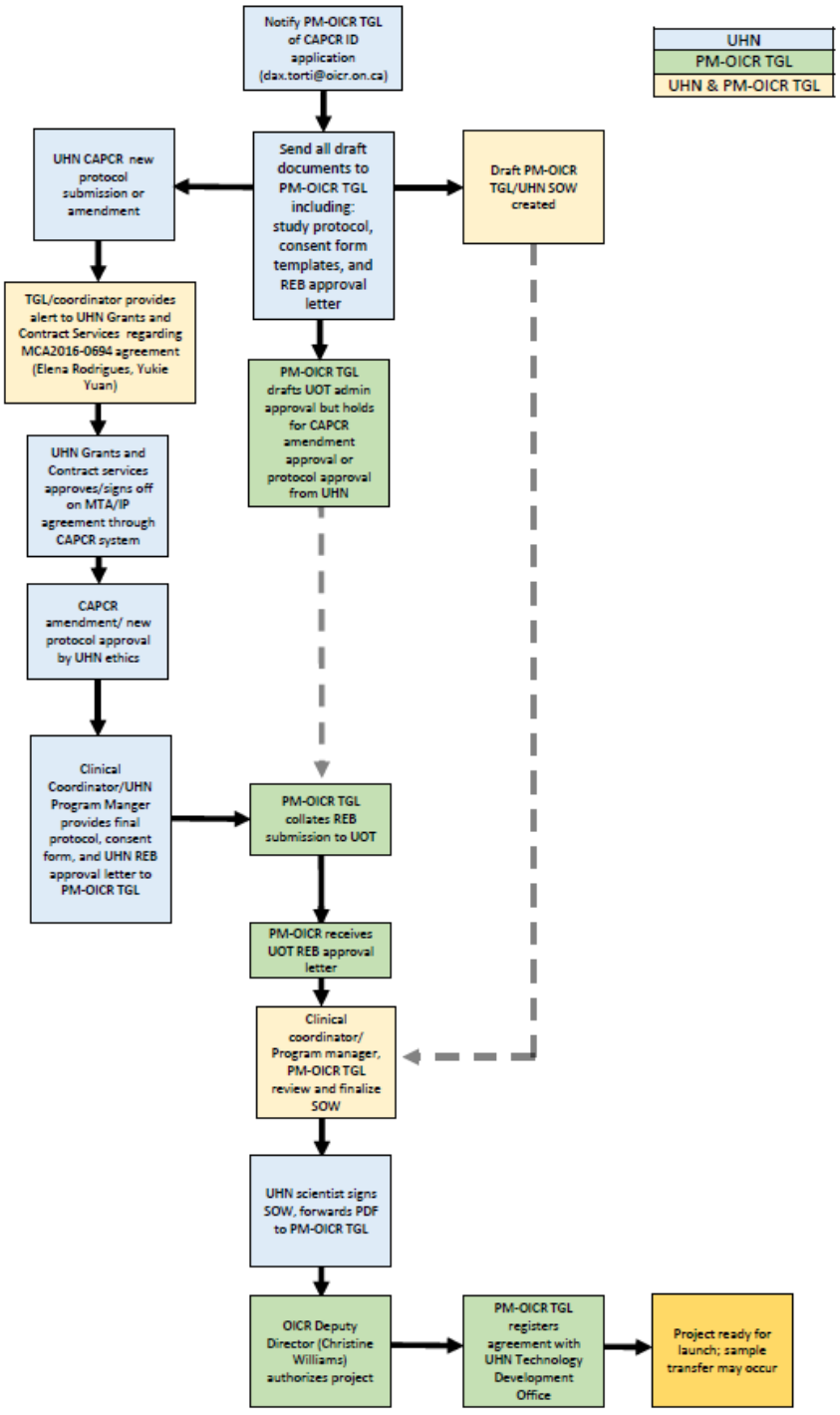
4.4 Infinium Methylation EPIC Array Informatics:

IDAT files generated from the Illumina iScan array scanner will be preprocessed using Bioconductor package minfi 1.2 [21]. Data will be normalized wth ssNoob. A correction for tissue type (FFPE/frozen) may be performed by using the removeBatchEffect function of the Limma package version 3.30.9 [22]. M-values will be calculated based on log 2 ratios of the intensities of methylated versus unmethylated probes prior to annotation. Signature discovery may use a combination of Spectral Clustering (SNF), Multidimensional Scaling Plot (MDS), Hierarchical clustering, consensus clustering, or other methods to achieve a stable clustering of a signature probe set.

5. Contracts, MTAs and DTAs

PM-OICR TGL is a joint collaborative initiative co-funded by OICR and UHN. UHN collaborations with PM-OICR TGL are covered under the master agreement MCA 2016-0694. This agreement allows for a streamlined workflows, and rapid project initiation. This contract includes all terms relevant to materials/data transfer and intellectual property. UHN collaborators are required to complete a statement of work form (available on our website) which is registered at UHN's Technology

Development and Commercialization (TDC) office. These agreements require iRB and University of Toronto REB reviews and/or amendments. Please see the workflow diagram below and/or contact Dax Torti (dax.torti@oicr.on.ca) for assistance. Collaborations outside of UHN are covered by separate agreements.



PM-OICR TGL iRB and Contracts Workflow Diagram

6. Revision History

Version Number	Date *(yyyy-mm-dd)*	History of change
1.0	2017-08-18	Template document created (DT)

7. References:

1. McKenna, A., et al., *The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data.* Genome Res, 2010. **20**(9): p. 1297-303.
2. Cibulskis, K., et al., *Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples.* Nat Biotechnol, 2013. **31**(3): p. 213-9.
3. Li, H., et al., *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-9.
4. Saunders, C.T., et al., *Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs.* Bioinformatics, 2012. **28**(14): p. 1811-7.
5. McLaren, W., et al., *The Ensembl Variant Effect Predictor.* Genome Biol, 2016. **17**(1): p. 122.
6. Lek, M., et al., *Analysis of protein-coding genetic variation in 60,706 humans.* Nature, 2016. **536**(7616): p. 285-91.
7. Chang, M.T., et al., *Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity.* Nat Biotechnol, 2016. **34**(2): p. 155-63.
8. Memorial Sloan Kettering Cancer Centre, Q.D. *OncoKB*. 2016; Available from: Oncokb.org.
9. Landrum, M.J., et al., *ClinVar: public archive of interpretations of clinically relevant variants.* Nucleic Acids Res, 2016. **44**(D1): p. D862-8.
10. Favero, F., et al., *Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data.* Ann Oncol, 2015. **26**(1): p. 64-70.
11. John-William Sidhom, C.A.B., Jonathan J. Havel, Timothy A. Chan, Jonathan P. Schneck. *ImmunoMap: a novel bioinformatics tool for immune cell repertoire analysis*. in *AACR Annual Meeting 2017*. Washinton, DC.
12. Newman, A.M., et al., *Robust enumeration of cell subsets from tissue expression profiles.* Nat Methods, 2015. **12**(5): p. 453-7.
13. Cerami, E., et al., *The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data.* Cancer Discov, 2012. **2**(5): p. 401-4.
14. Gao, J., et al., *Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal.* Sci Signal, 2013. **6**(269): p. pl1.
15. Dobin, A., et al., *STAR: ultrafast universal RNA-seq aligner.* Bioinformatics, 2013. **29**(1): p. 15-21.
16. Li, B. and C.N. Dewey, *RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome.* BMC Bioinformatics, 2011. **12**: p. 323.
17. Yoshihara, K., et al., *Inferring tumour purity and stromal and immune cell admixture from expression data.* Nat Commun, 2013. **4**: p. 2612.

18. Subramanian, A., et al., *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.* Proc Natl Acad Sci U S A, 2005. **102**(43): p. 15545-50.
19. Haas, B., et al., *STAR-Fusion: Fast and Accurate Fusion Transcript Detection from RNA-Seq.* bioRxiv, 2017.
20. Krueger, F. and S.R. Andrews, *Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications.* Bioinformatics, 2011. **27**(11): p. 1571-2.
21. Fortin, J.P., T.J. Triche, Jr., and K.D. Hansen, *Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi.* Bioinformatics, 2017. **33**(4): p. 558-560.
22. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing and microarray studies.* Nucleic Acids Res, 2015. **43**(7): p. e47.