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 (UHN) 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;
- Facilitate rapid dissemination of patient tumor genomic reports and annotation through cbioportal, unifying genomic interpretation, pathology review and imaging into a common accessible database

2. Pathology/Correlative Lab Requisition Guidance

PM-OICR TGL recommends the following parameters for tissue requests: 10 micron sections, minimum n=10, 1 Hematoxylin and eosin (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. Please notify the consulting pathologist to circle tumor regions on H&E slides and the section's % total tumor area relative to normal tissue, and within the tumor area, the % viable tumor and % necrotic tumor for the circled region. Please provide a de-identified pathology report containing this information to TGL including tissue site and diagnosis. If H&E images are scanned at your institute, we request a minimum of 20X resolution (TGL scans at 40X) including pathology markup. Please forward image files as ".SVS" format (Aperio Image File Format, type code 3 recommended) via secure file transfer; images will be incorporated into cbioportal (Aperio eSlide manager interface). TGL prefers to receive marked H&E, but will accept digital scans with tumor markup to guide tissue extraction. If the

entire tissue section is tumor, the entire tissue may be circled and scored. Ideally a 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. PM-OICR TGL requests de-identified pathology reports be forwarded through secure file transfer.

3. Recommended Core Clinical Data Elements

PM-OICR TGL recommends the capture of core data elements endorsed by the Centre for Medical Technology Policy and Molecular Evidence Development Consortium [1]. When possible, it is advised to collect the following data elements:

- Demographics: gender, ethnicity, race, cause of death
- Medical history: prior malignancies
- Physical exam at first diagnosis: height, weight, performance status
- First diagnosis of cancer of interest: basis of diagnosis, cancer site and histology, stage and grade, site and type of tissue sampling, prognostic biomarkers (presence/absence and levels), additional molecular diagnostic testing and performing laboratory
- Treatment episode: therapeutic agent and/or modality, intent of treatment, reason for ending treatment
- Outcomes (for each assessment episode): disease response, method of response evaluation, sites of any progression/recurrence, vital status, performance status and weight
- Dates: year of birth, date of death, date of diagnosis of any prior malignancies, date of physical evaluation at diagnosis, date of definitive diagnosis, beginning and ending dates of treatment (for each treatment), date of assessment of outcome (for each assessment)

4. Shipment and/ or Sample Receipt at PM-OICR TGL

Materials may be transferred to PM-OICR TGL from UHN's Correlative Studies Program or UHN BioBank from the following addresses, or other sites as designated in study protocol:

Correlative Studies Program
c/o Vanessa Spears
Princess Margaret Cancer Centre
610 University Avenue, Suite 7-409

Toronto, Ontario
M5G0A3
416.946.4501 X2562
Vanessa.speers@uhn.ca

UHN BioBioBank
Toronto General Hospital, Eaton Wing
11th floor, RM1126
200 Elisabeth Street
Toronto, Ontario
M5G2C4
416.340.4800 X4744
Biospecifmen.sciences.program@uhn.ca

to PM-OICR TGL in person, or via scheduled pickup:

PM-OICR TGL
c/o Dax Torti
Ontario Institute for Cancer Research
661 University Avenue, Suite 510
Toronto, Ontario
M5G0A3
Dax.torti@oicr.on.ca
647.260.7938

Perishable materials including fresh frozen tissues, blood and derivatives (plasma, peripheral blood mononuclear cells (PBMCs), circulating free DNA (cfDNA) and circulating tumor DNA (ctDNA) must be shipped/packed on dry ice. Formalin fixed paraffin embedded (FFPE) tissues, blocks and/or sections may be shipped/packed in slide boxes at room temperature for transport.

All materials must contain study codes only (de-identified), and be accompanied with a sample manifest (sample submission sheet) available online at https://labs.oicr.on.ca/translational-genomics-laboratory/. Completed sample submission sheets must be emailed to dax.torti@oicr.on.ca and tglsamples@lists.oicr.on.ca prior to sample receipt.

All materials and data sent to PM-OICR TGL must be de-identified, removing any reference to direct identifiers including patient name, medical record number (MRN), chart number, and surgical pathology numbers.

5. Lab Protocols

5.1 Plasma and Buffy Coat Separation from Whole Blood Protocol

Up to four 10 ml tubes of whole blood collected in EDTA, STRECK Cell-Free DNA BCT, or ROCHE cell free DNA collection tubes may be processed for buffy coat and plasma. After thorough inversion of blood collection tube, transfer whole blood to new 15ml conical tube and spin at 1900g at 4°C for 10 minutes. Remove tube from centrifuge and transfer plasma layer to a new 15ml conical tube. Plasma layers from two whole blood collection tubes may be combined into one conical tube. After collection of the plasma layer, carefully pipette the buffy coat layer into 1.5ml tubes without disturbing the erythrocyte layer. Continue with processing the plasma layer by spinning the primary plasma isolate at 16,000g for 10 minutes. Transfer the isolated plasma to a new tube without disturbing the pellet of cellular debris leaving 0.5ml behind. Immediately freeze buffy coat and purified plasma at -80°C, or in liquid phase nitrogen for long term storage. PM-OICR TGL prefers plasma in 15 ml conical tubes, but will accept 1.5 ml aliquots for processing; ideally 10 mls of plasma are required for cfDNA purification. Recommended protocols are available on our website.

5.2 DNA Isolation from Buffy Coat Protocol

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

5.3 Circulating Tumor DNA (ctDNA) Isolation from Plasma Protocol

ctDNA is isolated using the Qiagen QIAamp Circulating Nucleic Acid Kit according to manufacturer's protocol. Briefly plasma is spun at 16,000g to remove residual cellular debris. Plasma is treated with proteinase K and digested at 60°C for 1 hour. Lysate is processed through QIAamp mini column using a vacuum manifold, and washed successively with kit wash buffers, prior to DNA elution.

5.4 DNA/RNA Co-isolation from Fresh Frozen Tumor Tissue Protocol

DNA and RNA are co-isolated from a maximum of 30 mg of frozen tissue. The mortar is filled with liquid nitrogen to pre-chill the mortar and pestle. A maximum of 30mg of tissue is placed in the chilled mortar and a small amount of liquid nitrogen is ladled into the mortar. Tissue is ground with a pestle to a fine powder, prior to tissue homogenization using Qiagen Qiashredder columns. DNA and RNA are co-isolated using the Qiagen AllPrep Mini kit according to manufacturer's guidelines.

5.5 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 FFPE 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.

5.6 RNA Sequencing Lab Protocol

RNA libraries are synthesized from 200 ng of Total RNA using the Illumina TruSeq Stranded Total RNA with Ribozero Gold Sample Prep kit. Total RNA is depleted of ribosomal RNA (including mitochrondrial ribosomal RNA), first and second strand cDNA is synthesized, A-tailed, adapter ligated, and PCR amplified. 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. Sequencer selection may vary depending on project specifications. All libraries are validated on the MiSeq platform prior to deep sequencing.

5.7 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 Illumina NovaSeq, HiSeq2500 or NextSeq550 platforms, V1, V4, and V2 chemistry and reagents respectively. Sequencer selection may vary depending on project specifications. All libraries are validated on the MiSeq platform prior to deep sequencing.

5.8 Low Pass (Shallow) Whole Genome Sequencing Lab Protocol

Whole genome libraries are prepared using the Lucigen NxSeq AmpFREE Low DNA Library kit according to manufacturer's directions. Briefly, 100 ng of DNA is sheared, prior to end repair, A-tailing, adapter ligation, and PCR amplification. Libraries are sequenced to a depth of 0.1X on the Illumina HiSeq2500 platform, V4 Chemistry and reagents, with read lengths of 125bpX125bp, or on the NextSeq550 platform, V2 Chemistry and reagents, with read lengths of 75bpX75bp. Sequencer selection may vary depending on project specifications.

5.9 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 may be sequenced on Illumina NovaSeq, HiSeq2500 or NextSeq550 platforms, V1, V4, and V2 chemistry and reagents respectively. Sequencer selection may vary depending on project specifications.

5.10 Methylation Array Protocol

Illumina Infinium Methylation EPIC BeadChip arrays are processed following manufacturer's instructions. 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.

5.11 cfMeDIP Cell Free Methylated DNA Immunoprecipitation (IP) Protocol

Capture libraries are prepared from a minimum of 10 ng of circulating free (cf) DNA isolated from plasma (see protocol 3.1). Pre-capture libraries are synthesized using a modified KAPA Hyper Prep Kit protocol. cfDNA is end repaired, A-tailed and adapter ligated (unique molecular index, UMI). The cfDNA library is spiked with methylated and unmethylated Arabidopsis DNA control and split into two processing streams; one for 5-methylcytosine IP and one as an input control. 5-methylcytosine antibody (Diagenode Mag MeDIP kit) is used to selectively enrich cfDNA by immunoprecipitation of the IP sample via a modified manufacturer's protocol, prior to library amplification and sequencing. Libraries may be sequenced on Illumina NovaSeq, HiSeq2500 or NextSeq550 platforms, V1, V4, and V2 Chemistry and reagents respectively to read depth of 50 million clusters, or 50 million reads. Sequencer selection may vary depending on project specifications.

5.12 CtDNA Targeted Sequencing Panels

ctDNA libraries are prepared from a minimum of 10 ng of circulating free (cf) DNA isolated from plasma (see protocol 3.1). Pre-capture libraries are synthesized using a modified protocol based on the KAPA Hyper Prep Kit, prior to capture using Integrated DNA Technology (IDT) gene probe sets. Briefly, DNA is end repaired, A-tailed, adapter ligated (unique molecular index, UMI), and PCR amplified. Gene specific probes are hybridized to DNA libraries for 4 hours and washed, prior to on-bead amplification and cleanup, following IDT's Hybridization capture of DNA libraries using xGEN Lockdown Probes and Reagents protocol, version 2. Captured ctDNA is sequenced to a depth of 10,000-20,000X and may be sequenced on Illumina NovaSeq, HiSeq2500 or NextSeq550 platforms, V1, V4, and V2 Chemistry and reagents respectively. Sequencer selection may vary depending on project specifications.

6. Informatics

6.1 Pre-analytic Library Validation Informatics

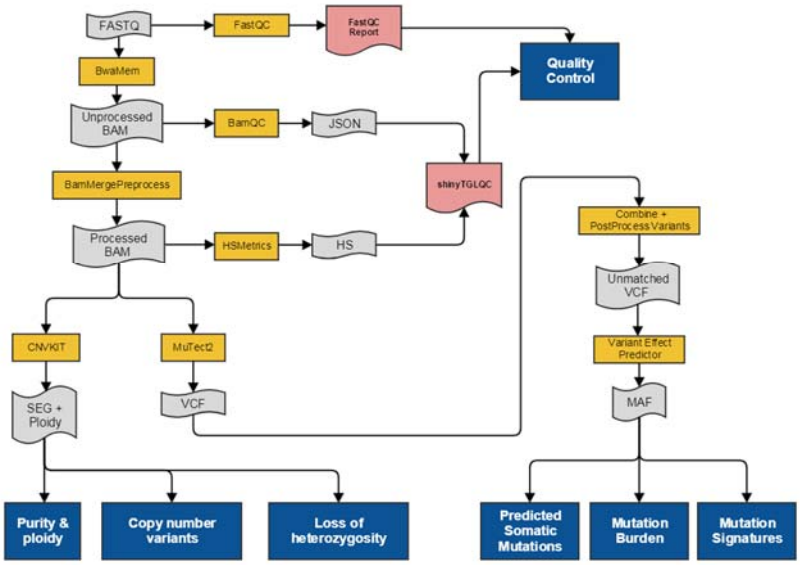
All sequencing libraries including exome and Total RNA libraries are validated for quality prior to deep sequencing. PM-OICR TGL performs pre-analytic sequencing on the MiSeq platform sequencing each library to a minimum read depth of 10,000 clusters, 150bpX8bpX8bpX150bp. Total RNA libraries are evaluated for ribosomal contaminant levels using RSeQC v2.6.4 [2]. FFPE or fresh frozen samples with ribosomal contaminant levels in excess of 35% of reads are failed and queued for repeat library synthesis. Exome libraries are validated for insert size and duplicate levels using FastQC [3]. Libraries with insert sizes <150bp are failed, or queued for repeat library synthesis. Additional quality control

metrics generated by FastQC are captured in our quality control database, Shiny TGLQC, for review of sequencing library quality by lab technicians.

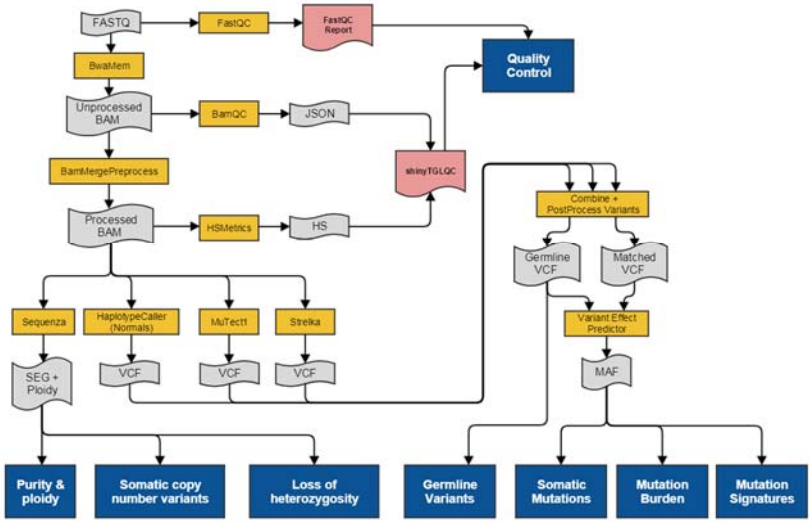
6.2 Exome Informatics

Sequence reads are analyzed with FastQC [3] and aligned against hg19 human reference using BwaMem v 0.7.12 to generate raw sequence alignments in BAM format [4]. Preprocessing, which includes PCR-duplicate marking, indel re-alignment and base quality recalibration is performed using Picard v1.72 and GATK v3.6.0 [5]. Quality control metrics are captured within our quality control database, Shiny TGLQC. When available, preprocessing will be performed in matched tumor/normal pairs to improve indel re-alignment (see figure 6.2b). Haplotype Caller [6], MuTect1 v1.1.7 [7] and Strelka v1.0.13 [8] will be run to create raw variant call files (VCFs).

Raw VCF files will be annotated with Variant Effect Predictor v92 [9] against multiple variant annotation databases. Germline and somatic variants will be annotated with GnomAD allele frequencies [10] – a database of variant frequencies in a healthy population – in order to remove common variants. Variants will also be annotated against known cancer hotspots (CancerHotspots.org) both at the variant level and gene level [10, 11]. Analysis will include actionable /oncogenic driver analysis using the Precision Oncology Knowledge Base (oncoKb) and pathogenic database ClinVar [12, 13] . Global mutation signatures defined in Stratton et al. [14] will also be assessed using deconstructSigs v1.8.0 [15]. 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, loss of heterozygosity, and to estimate ploidy/cellularity using Sequenza for matched tumor/normal pairs [16]. In the absence of matched blood (buffy coat)/normal tissue, a pooled reference is constructed from aggregate blood (buffy coat) samples and used to normalize depth of coverage in tumor samples for CNV analysis using CNVKIT v0.9.1 [17]. Mutation burden will be calculated as the number of non-synonymous mutations per callable megabase; MuTect v1.1.7 [7] wig coverage file is used to determine callability. 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 [18, 19] for visualization of genomic data cohorts.



(a) Figure 6.2a: Tumor only exome pipeline.



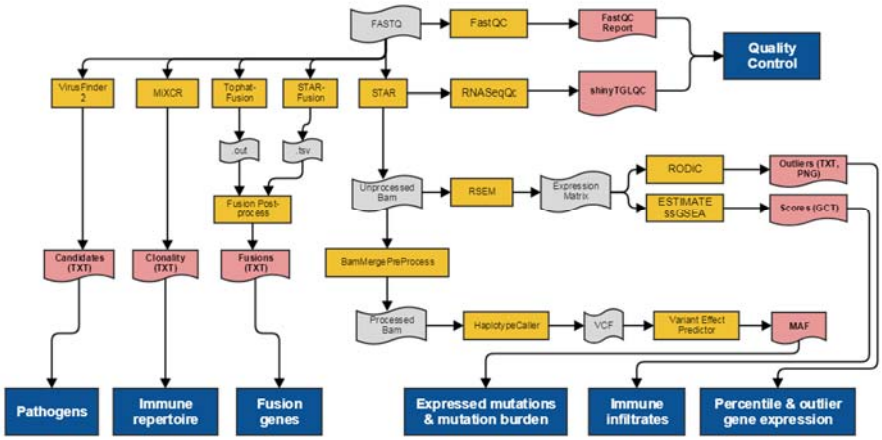
(b) Figure 6.2b: Tumor- normal (paired) exome pipeline.

6.3 Low Pass (Shallow) Whole Genome Informatics

Reads are aligned using BwaMem v0.7.9a to generate raw sequence alignments in BAM format [4]. Aligned sequences are pre-processed using GATK v3.5 and Picard v1.9.1; pre-processing includes PCR duplicate marking, indel re-alignment in matched tumour-normal pairs and base recalibration [5]. Picard CollectWgsMetrics and CollectInsertSizeMetrics tools are used for quality control. Depth of coverage is estimated using Bedtools v2.23.0 [20]. Copy number profiling is performed using QDNAseq v1.14.0 using a bin size of 50 kb [21]. Loss of heterozygosity is estimated by comparing the heterozygous single-nucleotide polymorphism distribution profiles generated from Mutect v1.1.5 [7] against a reference SNP distribution.

6.4 RNA Informatics

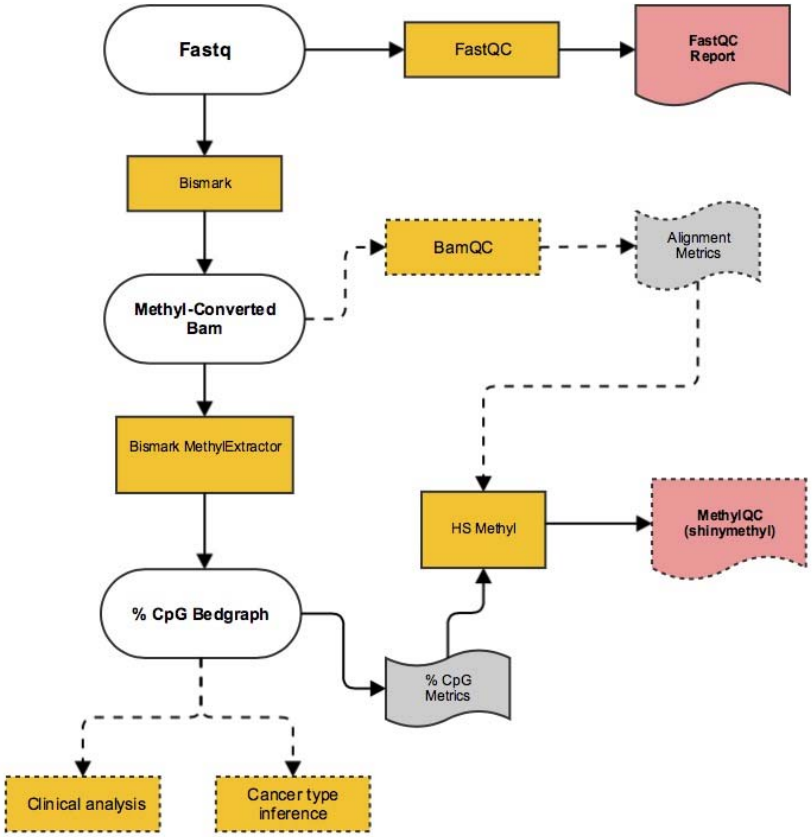
PM-OICR TGL will process Total RNA through FastQC [3], STAR aligner v2.6.0c [22], ReSeQC v2.6.4 [3] followed by RNA abundance quantification using RSEM v1.3.0 [23] to generate an expression matrix used for expression outlier analysis using RNA-seq Outlier Detection in Cancer (RODIC)[24], ESTIMATE [25] for immunological gene signatures/infiltrates and ssGSEA [26] for pathway analysis. Quality control metrics are captured in Shiny TGLQC and reviewed. 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. HaplotypeCaller [6]will generate variant call files (VCF), prior to variant effect predictor analysis of annotated mutations [5]. Tophat Fusion Detection v2.0.10 [27] and STAR-fusion v1.4.0 [28] will be run for detection of fusion candidates. MiXCR v2.1.10 [29] will be applied for analysis of tumor T and B cell receptor repertoire, and tumor tissues assessed for expressed viral transcripts using VirusFinder 2 [30]. 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 [18, 19] for visualization of genomic data cohorts.



(a) Figure 6.4a: Tumor transcriptome pipeline.

6.5 Methylation Sequencing Informatics

Sequence reads from Illumina TruSeq Methylation EPIC exome libraries will be aligned using Bismark software v0.15.0 [31]. 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 [18, 19] for visualization of genomic data cohorts.



(a) Figure 6.5a: Tumor methylation pipeline.

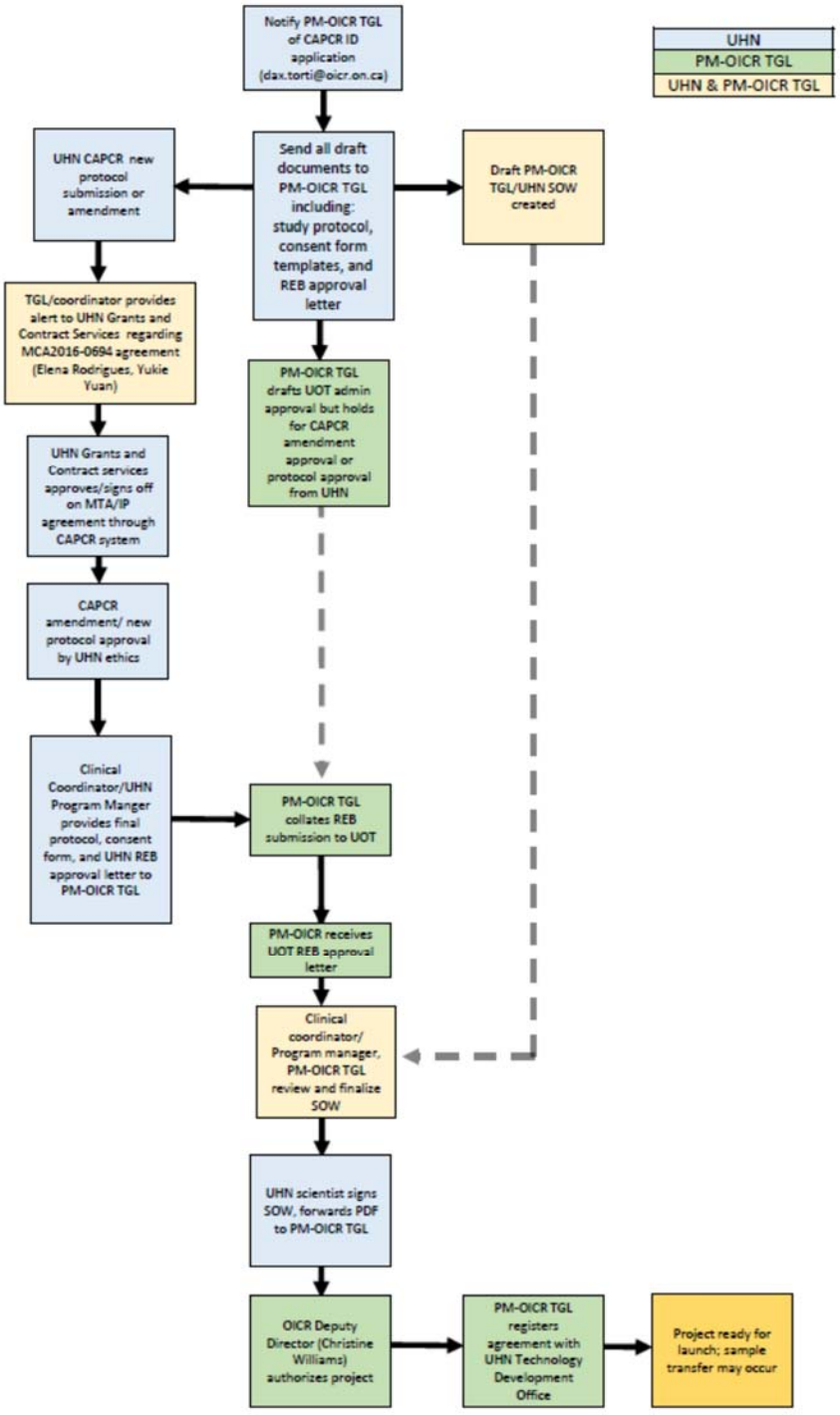
6.6 Infinium Methylation EPIC Array Informatics

IDAT files generated from the Illumina iScan array scanner will be preprocessed using Bioconductor package minfi 1.2 [32]. Data will be normalized with ssNoob. A correction for tissue type (FFPE/frozen)

may be performed by using the removeBatchEffect function of the Limma package version 3.30.9 [33]. 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.

7. 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 or with other OICR workgroups are covered by separate/independent agreements.



(a) Figure 7a: PM-OICR TGL iRB and Contracts Workflow Diagram

8. Revision History

Version Number	Date *(yyyy-mm-dd)*	History of change
1.0	2017-08-18	Template document created (DT)
1.1	2017-08-30	Added buffy coat/plasma isolation, ctDNA isolation, low pass whole genome library prep
1.2	2017-09-08	Updated Cibersort/Immunomap information
1.3	2018-02-16	Added protocols for cfMeDIP, ctDNA. Added Illumina NovaSeq platform to protocols. (Dax Torti)
1.4	2018-02-27	Updated exome informatics with tumor only methods including CNV kit, and global mutational signatures method with deconstructSigs. Removed certain tools from RNA Informatics. Updated references. (Jon Torchia/Dax Torti)
1.5	2018-03-20	Included shipping address and procedures, added low pass whole genome informatics process (Dax Torti)
1.6	2018-07-04	Added recommended core clinical data elements and added guidance on de-identification of patient direct identifiers, updated exome and transcriptome informatic pipeline documentation (Dax Torti)

9. References:

1. Conley, R.B., et al., *Core Clinical Data Elements for Cancer Genomic Repositories: A Multi-stakeholder Consensus.* Cell, 2017. **171**(5): p. 982-986.
2. Wang, L., S. Wang, and W. Li, *RSeQC: quality control of RNA-seq experiments.* Bioinformatics, 2012. **28**(16): p. 2184-5.
3. Andrews, S., *FastQC.* 2018.
4. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform.* Bioinformatics, 2009. **25**(14): p. 1754-60.
5. 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.
6. Poplin, R., et al., *Scaling accurate genetic variant discovery to tens of thousands of samples.* bioRxiv, 2017.
7. Cibulskis, K., et al., *Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples.* Nat Biotechnol, 2013. **31**(3): p. 213-9.
8. Saunders, C.T., et al., *Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs.* Bioinformatics, 2012. **28**(14): p. 1811-7.
9. McLaren, W., et al., *The Ensembl Variant Effect Predictor.* Genome Biol, 2016. **17**(1): p. 122.

10. Lek, M., et al., *Analysis of protein-coding genetic variation in 60,706 humans.* Nature, 2016. **536**(7616): p. 285-91.
11. 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.
12. Memorial Sloan Kettering Cancer Centre, Q.D. *OncoKB*. 2016; Available from: Oncokb.org.
13. Landrum, M.J., et al., *ClinVar: public archive of interpretations of clinically relevant variants.* Nucleic Acids Res, 2016. **44**(D1): p. D862-8.
14. Alexandrov, L.B., et al., *Signatures of mutational processes in human cancer.* Nature, 2013. **500**(7463): p. 415-21.
15. Rosenthal, R., et al., *DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution.* Genome Biol, 2016. **17**: p. 31.
16. Favero, F., et al., *Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data.* Ann Oncol, 2015. **26**(1): p. 64-70.
17. Talevich, E., et al., *CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing.* PLoS Comput Biol, 2016. **12**(4): p. e1004873.
18. 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.
19. Gao, J., et al., *Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal.* Sci Signal, 2013. **6**(269): p. pl1.
20. Quinlan, A.R. and I.M. Hall, *BEDTools: a flexible suite of utilities for comparing genomic features.* Bioinformatics, 2010. **26**(6): p. 841-2.
21. Scheinin, I., et al., *DNA copy number analysis of fresh and formalin-fixed specimens by shallow whole-genome sequencing with identification and exclusion of problematic regions in the genome assembly.* Genome Res, 2014. **24**(12): p. 2022-32.
22. Dobin, A., et al., *STAR: ultrafast universal RNA-seq aligner.* Bioinformatics, 2013. **29**(1): p. 15-21.
23. 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.
24. Torchia, J., *RNA Seq Outlier Detection (RODIC)*. 2018.
25. Yoshihara, K., et al., *Inferring tumour purity and stromal and immune cell admixture from expression data.* Nat Commun, 2013. **4**: p. 2612.
26. 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.
27. Kim, D. and S.L. Salzberg, *TopHat-Fusion: an algorithm for discovery of novel fusion transcripts.* Genome Biol, 2011. **12**(8): p. R72.
28. Haas, B., et al., *STAR-Fusion: Fast and Accurate Fusion Transcript Detection from RNA-Seq.* bioRxiv, 2017.
29. Bolotin, D.A., et al., *MiXCR: software for comprehensive adaptive immunity profiling.* Nature Methods, 2015. **12**: p. 380.
30. Wang, Q., P. Jia, and Z. Zhao, *VERSE: a novel approach to detect virus integration in host genomes through reference genome customization.* Genome Medicine, 2015. **7**(1): p. 2.
31. Krueger, F. and S.R. Andrews, *Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications.* Bioinformatics, 2011. **27**(11): p. 1571-2.
32. 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.

33. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing and microarray studies.* Nucleic Acids Res, 2015. **43**(7): p. e47.