RESEARCH

Assessment of a Clinical Targeted Sequencing Panel for Germline Pharmacogenomic Testing in Formalin-fixed Paraffin-embedded Tumours

Shyong Quin Yap[1] and Aly Karsan[1,2*]

*Correspondence:
akarsan@bcgsc.ca
[1]British Columbia Cancer Research Centre, 675 West 10th Ave, V5Z 1L3 Vancouver, BC, Canada
Full list of author information is available at the end of the article

Abstract

Background: Because the tumour genome contains germline information, clinical laboratories can leverage tumour genomic testing to perform initial screening for clinically relevant germline variants such as variants in pharmacogenomic (PGx) genes. Subsequently, only patients with potential germline PGx variants would be subjected to downstream germline testing. While this procedure is more cost-effective than sequencing and analyzing normal DNA for every patient, the concordance of germline PGx variants between tumour and matched normal specimens must be evaluated to determine the feasibility of detecting germline PGx variants in tumour DNA.

Results:

Conclusions:

Keywords: Tumour sequencing; Targeted next-generation sequencing panel; Germline pharmacogenomics testing; Formalin-fixed paraffin-embedded tumours

Background

Tumour profiling using next-generation sequencing (NGS) technologies has been increasingly integrated into oncologic care to detect targetable somatic mutations and personalize treatments for cancer patients. Although analysis of tumour-normal paired samples is required to accurately discriminate between somatic and germline variants, most clinical laboratories only sequence tumour samples to minimize cost and turnaround time [1]. However, genomic analyses of tumours can also reveal secondary genomic findings, which are germline information that may have clinical implications for patients and their family members [1]. In fact, several studies demonstrated that a germline cancer-predisposing variant is present in 3-10% of patients undergoing tumour-normal sequencing [1, 2, 3, 4]. Therefore, clinical laboratories providing tumour genomic testing must be equipped to perform germline confirmatory testing on potential germline variants or be prepared to refer such cases to external services.

Because the tumour genome contains germline information, clinical laboratories can leverage tumour genomic testing to perform initial screening for clinically relevant germline variants such as variants in pharmacogenomic (PGx) genes. Subsequently, a similar framework for validating secondary germline findings can be applied, in which only patients with potential germline PGx variants are subjected to downstream germline testing. This procedure for germline PGx testing is more

cost-effective because it does not require processing, sequencing, and analysis of normal DNA for every patient. The ability to implement germline PGx testing at a reduced cost can significantly benefit patient care because these variants cause functional changes in drug targets and drug disposition proteins (proteins involved drug metabolism and transport), thereby contributing to inter-patient differences in chemotherapeutic response [5]. Hence, such genomic information can be used to guide the selection of chemotherapeutic drugs and optimization of drug dosage for cancer patients, leading to improved safety and efficacy of treatment and reduced risk of toxicity [5].

Detection of genomic alterations in tumour DNA is also faced with technical challenges conferred by formalin-fixed paraffin-embedded (FFPE) tumour specimens [6, 7]. Tumour biopsies are often formalin-fixed to preserve tissue morphology for histological examination and to enable storage at room temperature; however, formalin fixation causes DNA fragmentation and base modifications, which pose difficulties in using DNA extracted from FFPE tumours for clinical genomic testing [6, 7]. Fragmentation damage caused by formalin fixation leads to reduced template DNA for PCR amplification, thereby affecting the efficiency of amplicon-based NGS testing [6, 7]. Furthermore, the degree of DNA fragmentation was shown to be higher in tissues from older FFPE blocks and tissues fixed with formalin of lower pH [6]. Formalin fixation is also problematic because it gives rise to depurination, which generates abasic sites, and cytosine deamination resulting in C>T/G>A transitions [6]. These forms of formalin-induced DNA damage contribute to the presence of sequence artifacts in FFPE specimens, which can be inaccurately identified as real genomic alterations.

In this study, we assessed the concordance of germline PGx variants between tumour and matched normal DNA by analyzing amplicon-based targeted NGS data from 213 patients with tumour-normal paired samples. While matched normal DNA was derived from peripheral blood, tumour DNA was extracted from FFPE tumour blocks; thus, we compared the quality metrics of sequencing data between tumour and blood specimens and evaluated the prevalence of formalin-induced DNA damages to address the impact of formalin fixation on amplicon-based NGS testing. We demonstrated that germline PGx variants can be identified with high sensitivity and precision in FFPE tumour DNA using a clinical targeted sequencing panel.

Methods

Patient Samples

Blood and FFPE tumour specimens were acquired from 213 patients who provided informed consent for The OncoPanel Pilot (TOP) study, a pilot study to optimize the OncoPanel, which is an amplicon-based targeted NGS panel for solid tumours, and assess its application for guiding disease management and therapeutic intervention. Patients in the TOP study are those with advanced cancers including colorectal cancer, lung cancer, melanoma, gastrointestinal stromal tumor (GIST), and other cancers (Table 1). The age of paraffin block for tumour specimens ranges from 18 to 5356 days with a median of 274 days.

OncoPanel (Solid Tumour NGS Panel)

The OncoPanel is offered by the British Columbia Cancer Agency (BCCA) for clinical genomic testing of coding exons and clinically relevant hotpots of 20 cancer-related genes and six PGx genes: *DPYD*, *GSTP1*, *MTHFR*, *TYMP*, *TYMS*, and *UGT1A1*. Full list of genes and gene reference models for the OncoPanel is presented in Table 3. Primers were designed by RainDance Technologies (Lexington, MA) using the GRCh37/hg19 reference sequence to generate 416 amplicons between 100 bp and 250 bp in size, which interrogate $\sim$ 20 kb of target bases. Target regions of the six PGx genes in the OncoPanel were assayed by 49/416 amplicons. Complete lists of primers and amplicons are provided in the Supplemental Materials.

Sample Preparation, Library Construction, and Illumina Sequencing

Genomic DNA was extracted from blood and FFPE tumour specimens using the Gentra Autopure LS DNA preparation platform and QIAamp DNA FFPE tissue kit (Qiagen, Hilden, Germany) respectively. The extracted DNA was sheared according to a previously described protocol [8] to attain approximate sizes of 3 kb followed by PCR primer merging, amplification of target regions, and adapter ligation using the Thunderstorm NGS Targeted Enrichment System (RainDance Technologies, Lexington, MA) as per manufacturer's protocol. Barcoded amplicons were sequenced with the Illumina MiSeq system for paired end sequencing with a v2 250-bp kit (Illumina, San Diego, CA).

Variant Calling Pipeline

Reads that passed the Illumina Chastity filter were aligned to the GRCh37/hg19 human reference genome using the BWA mem algorithm (version 0.5.9) with default parameters, and the alignments were processed and converted to the BAM format using SAMtools (version 0.1.18). Variant calling was performed with the SAMtools `mpileup` function `(samtools mpileup -BA -d 500000 -L 500000 -q 1)` to generate pileup files for all target bases followed by the VarScan2 `mpileup2cns` (version 2.3.6) function with parameter thresholds of variant allele frequency $\geq$ 10% and Phred base quality $\geq$ 20 `(--min-var-freq 0.1 --p-value 0.01 --strand-filter 0 --output-vcf --variants --min-avg-qual 20)`. Variant calls were filtered using the VarScan2 `fpfilter` function with fraction of variant reads from each strand $\geq$ 0.1 and default thresholds for other parameters. SnpEff (version 4.2) was used for variant annotation and effect prediction whereas the SnpSift package in SnpEff was used to annotate variants with databases such as dbSNP (b138), COSMIC (version 70), 1000 Genomes Project, ClinVar, and ExAC (release 0.3) for interpretation.

Data Analysis

Specifics on how each analysis was performed.

Results

Some description of the workflow over here.

Comparison of Amplicon Generation, Sequence Alignment, and Coverage Statistics between FFPE Tumours and Blood

To examine the effects of formalin-induced DNA damages on molecular testing, we determined whether FFPE tumours yield similar NGS quality metrics to blood (Table 2). We compared the amount of amplicon generated between FFPE tumours and blood and found a significant difference in the distributions of amplicon yield between specimen types (Wilcoxon signed-rank test, W=23 613, $P<2.2 \times 10^{-16}$). Although the amount of DNA input for amplicon generation was also significantly different between FFPE tumours and blood (Wilcoxon signed-rank test, W=15004, $P=3.5 \times 10^{-4}$), Spearman correlation showed that amplicon yield was weakly correlated with DNA input for amplicon generation (blood, $r=0.29$, 95% CI=0.16–0.40, $P=2.1 \times 10^{-5}$; FFPE tumour, $r=0.25$, 95% CI=0.12–0.37, $P=2.5 \times 10^{-4}$). This result implies that specimen type affects the efficiency of amplicon generation with FFPE tumours demonstrating a lower median amplicon yield of 103.6 ng than the median amplicon yield of blood, which is 299.2 ng.

We also compared read alignment between FFPE tumours and blood to examine whether different specimen types result in comparable sequencing data. For the proportion of reads that aligned to target sequences used for variant calling, which are known as on-target aligned reads, we found no significant difference in distributions between FFPE tumours and blood (Wilcoxon signed-rank test, W=10 178.5, $P=9.4 \times 10^{-2}$). Similarly, no significant difference in the proportion of off-target aligned reads (i.e. reads that mapped to the human reference genome but not target sequences) was demonstrated between FFPE tumours and blood (Wilcoxon signed-rank test, W=11 494.5, $P=0.7$). However, differences in the proportions of unaligned and contaminant reads were significant between specimen types (Wilcoxon signed-rank test, unaligned: W=19 069, $P=5.2 \times 10^{-15}$; contaminant: W=14 877, $P=9.9 \times 10^{-4}$). While NGS libraries generated from FFPE tumour and blood DNA differed in proportions of unaligned and contaminant reads, their libraries resulted in comparable proportion of on-target aligned reads; hence, providing similar amount of usable sequencing data for variant calling.

To determine whether the differences in amplicon yield and sequence alignment affect the ability to obtain adequate coverage depth for target bases of the OncoPanel, we

determined the feasibility of detecting germline variants this will recompared several quality metrics such as efficiency of amplicon generation, read alignment, and coverage statistics we compared several Because formalin fixation causes DNA damage and fragmentation, we assessed the influence of formalin fixation on compared amplicon yield between blood and FFPE specimens.

evaluated the potential effects of using FFPE tumour for detection of germline variants by

Evaluation of Amplification Performance

Formalin-induced sequence artifacts are prevalent at low variant allele frequency

Germline PGx variants can be detected with high sensitivity and precision in FFPE tumours

Discordant germline PGx variants are caused by low coverage sites

Discussion

Conclusions

Competing interests

The authors declare that they have no competing interests.

Author's contributions

Text for this section . . .

Acknowledgements

Text for this section . . .

Author details

[1]British Columbia Cancer Research Centre, 675 West 10th Ave, V5Z 1L3 Vancouver, BC, Canada. [2]Department of Pathology and Laboratory Medicine, University of British Columbia, Random Street, Random Post Code Vancouver, BC, Canada.

References

1. Raymond, V.M., Gray, S.W., Roychowdhury, S., Joffe, S., Chinnaiyan, A.M., Parsons, D.W., Plon, S.E.: Germline findings in tumor-only sequencing: Points to consider for clinicians and laboratories. Journal of the National Cancer Institute **108**(4), 1–5 (2016). doi:10.1093/jnci/djv351
2. Meric-Bernstam, F., Brusco, L., Daniels, M., Wathoo, C., Bailey, A.M., Strong, L., Shaw, K., Lu, K., Qi, Y., Zhao, H., Lara-Guerra, H., Litton, J., Arun, B., Eterovic, A.K., Aytac, U., Routbort, M., Subbiah, V., Janku, F., Davies, M.A., Kopetz, S., Mendelsohn, J., Mills, G.B., Chen, K.: Incidental germline variants in 1000 advanced cancers on a prospective somatic genomic profiling protocol. Annals of Oncology **27**(5), 795–800 (2016). doi:10.1093/annonc/mdw018
3. Schrader, K.A., Cheng, D.T., Joseph, V., Prasad, M., Walsh, M., Zehir, A., Ni, A., Thomas, T., Benayed, R., Ashraf, A., Lincoln, A., Arcila, M., Stadler, Z., Solit, D., Hyman, D., Zhang, L., Klimstra, D., Ladanyi, M., Offit, K., Berger, M., Robson, M.: Germline Variants in Targeted Tumor Sequencing Using Matched Normal DNA. JAMA oncology **2**(1), 1–8 (2015). doi:10.1001/jamaoncol.2015.5208
4. Jones, S., Anagnostou, V., Lytle, K., Parpart-li, S., Nesselbush, M., Riley, D.R., Shukla, M., Chesnick, B., Kadan, M., Papp, E., Galens, K.G., Murphy, D., Zhang, T., Kann, L., Sausen, M., Angiuoli, S.V., Jr, L.A.D., Velculescu, V.E.: Personalized genomic analyses for cancer mutation discovery and interpretation. Science Translational Medicine **7**(283), 283–53 (2015). doi:10.1126/scitranslmed.aaa7161
5. McLeod, H.L.: Cancer Pharmacogenomics: Early Promise, But Concerted Effort Needed. Science **339**(March), 1563–1566 (2013). doi:10.1126/science.1234139
6. Do, H., Dobrovic, A.: Sequence artifacts in DNA from formalin-fixed tissues: Causes and strategies for minimization. Clinical Chemistry **61**(1), 64–71 (2015). doi:10.1373/clinchem.2014.223040
7. Wong, S.Q., Li, J., Y-C Tan, A., Vedururu, R., Pang, J.-M.B., Do, H., Ellul, J., Doig, K., Bell, A., MacArthur, G.A., Fox, S.B., Thomas, D.M., Fellowes, A., Parisot, J.P., Dobrovic, A.: Sequence artefacts in a prospective series of formalin-fixed tumours tested for mutations in hotspot regions by massively parallel sequencing. BMC Medical Genomics **7**(1), 1–10 (2014). doi:10.1186/1755-8794-7-23
8. Bosdet, I.E., Docking, T.R., Butterfield, Y.S., Mungall, A.J., Zeng, T., Coope, R.J., Yorida, E., Chow, K., Bala, M., Young, S.S., Hirst, M., Birol, I., Moore, R.A., Jones, S.J., Marra, M.A., Holt, R., Karsan, A.: A clinically validated diagnostic second-generation sequencing assay for detection of hereditary BRCA1 and BRCA2 mutations. Journal of Molecular Diagnostics **15**(6), 796–809 (2013). doi:10.1016/j.jmoldx.2013.07.004

Figures

Figure 1 Workflow for Library Construction and Sequencing. A short description of the figure content should go here.

Figure 2 Sample figure title. Figure legend text.

Tables

Table 1 Distribution of cancer types in the TOP cohort.

Cancer Type	Number of Cases	Percentage (%)
Colorectal	97	46
Lung	59	28
Melanoma	18	8
Other*	17	8
GIST	7	3
Sarcoma	4	2
Neuroendocrine	4	2
Cervical	2	0.9
Ovarian	2	0.9
Breast	2	0.9
Unknown	1	0.5

*This category includes thyroid, peritoneum, lung sarcomatoid carcinoma, Fallopian tube, gastric, endometrial, squamous cell carcinoma, anal, salivary gland, peritoneal epithelial mesothelioma, adenoid cystic carcinoma, pancreas, breast, gall bladder, parotid epithelial myoepithelial carcinoma, and small bowel cancers.

Table 2 Assessment of Amplicon Generation, Sequence Alignment, and Coverage Statistics.

	Blood		FFPE Tumour		
Parameter	Median	Range	Median	Range	*P* value
DNA Input (ng)	147.8	55.5–266.4	140.9	11.8–271.0	3.5×10^{-4}
Amplicon Yield (ng)	299.2	84.0–1438.0	103.6	11.6–325.5	2.2×10^{-16}
On-target Aligned Reads (%)	86.8	74.0–95.9	88.4	32.5–97.4	9.4×10^{-2}
Off-target Aligned Reads (%)	7.3	0.8-24.0	8.4	0.4–60.4	0.7
Unaligned Reads (%)	1.3	0.4–7.2	0.8	0.3–12.1	5.2×10^{-15}
Contaminant (%)	3.9	0.1–8.1	1.4	0.03–63.6	9.9×10^{-4}
Zero Coverage Target Bases (%)	0.00	0.00-0.01	0.00	0.00-0.36	–
$\geq$ 100x Target Bases (%)	100.00	99.97-100.00	100.00	54.95–100.00	0.58
$\geq$ 1000x Target Bases (%)	100.00	99.00-100.00	99.98	20.61-100.00	1.92×10^{-5}

Table 3 Gene Reference Models for Genes in the OncoPanel.

Gene	Protein	Reference Model
AKT1	Protein kinase B	NM_001014431.1
ALK	Anaplastic lymphoma receptor tyrosine kinase	NM_004304.3
BRAF	Serine/threonine-protein kinase B-Raf	NM_004333.4
DPYD	Dihydropyrimidine dehydrogenase	NM_000110.3
EGFR	Epidermal growth factor receptor	NM_005228.3
ERBB2	Receptor tyrosine-protein kinase erbB-2	NM_001005862.1
GSTP1	Glutathione S-rransferase pi 1	NM_000852.3
HRAS	GTPase HRas	NM_005343.2
IDH1	Isocitrate dehydrogenase 1	NM_005896.2
IDH2	Isocitrate dehydrogenase 2	NM_002168.2
KIT	Tyrosine-protein kinase Kit	NM_000222.2
KRAS	KRas proto-oncogene GTPase	NM_033360.2
MAPK1	Mitogen-activated protein kinase 1	NM_002745.4
MAP2K1	Mitogen-activated protein kinase kinase 1	NM_002755.3
MTHFR	Methylenetetrahydrofolate reductase	NM_005957.4
MTOR	Serine/threonine-protein kinase mTOR	NM_004958.3
NRAS	Neuroblastoma RAS viral oncogene homolog	NM_002524.3
PDGFRA	Platelet-derived growth factor receptor alpha	NM_006206.4
PIK3CA	Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha	NM_006218.2
PTEN	Phosphatase and tensin homolog	NM_000314.4
STAT1	Signal transducer and activator of transcription 1	NM_007315.3
STAT3	Signal transducer and activator of transcription 3	NM_139276.2
TP53	Tumor protein P53	NM_000546.5
TYMP	Thymidine phosphorylase	NM_001113755.2
TYMS	Thymidylate synthetase	NM_001071.2
UGT1A1	Uridine diphosphate (UDP)-glucuronosyl transferase 1A1	NM_000463.2

Additional Files

Additional file 1 — Sample additional file title
Additional file descriptions text (including details of how to view the file, if it is in a non-standard format or the file extension). This might refer to a multi-page table or a figure.

Additional file 2 — Sample additional file title
Additional file descriptions text.