

Germline Variant Calling in Formalin-fixed Paraffin-embedded Tumours

Eva Yap, MSc. Student

October 30, 2017

Overview

① Background

② Research Question

③ Project Aims

④ Aim 1

⑤ Aim 2

⑥ Aim 3

⑦ Conclusions

Overview

1 Background

2 Research Question

3 Project Aims

4 Aim 1

5 Aim 2

6 Aim 3

7 Conclusions

The era of precision oncology



BC Cancer Agency's OncoPanel

New genetic test will personalize cancer treatment for B.C. residents



G. MARION JOHNSON

[More from G. Marion Johnson](#)

Published on: April 18, 2017 | Last Updated: April 18, 2017 11:27 AM PDT

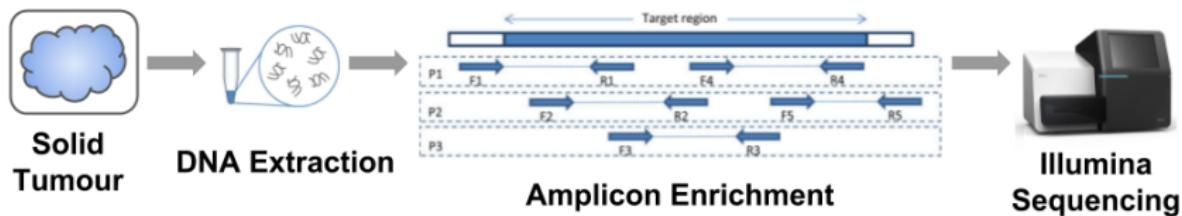
One of the greatest breakthroughs in human history, the Human Genome Project helped unlock the secrets of disease. By going back to the very building blocks of human beings, the future of modern medicine is in genetic testing. For cancer care in British Columbia, that future is now.



BC Cancer Agency's Dr. Hagen Kennecke helped develop the OncoPanel, which tests for genetic markers in cancer patients to help determine best treatments. *SUPPLIED*

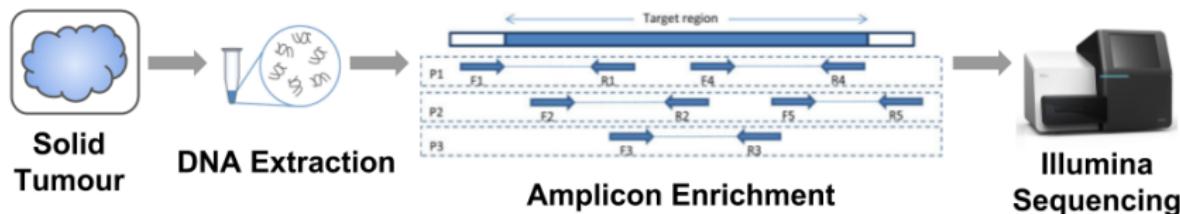
<http://vancouversun.com/uncategorized/staff/health-blogs/new-genetic-test-will-personalize-cancer-treatment-for-b-c-residents>

BC Cancer Agency's OncoPanel



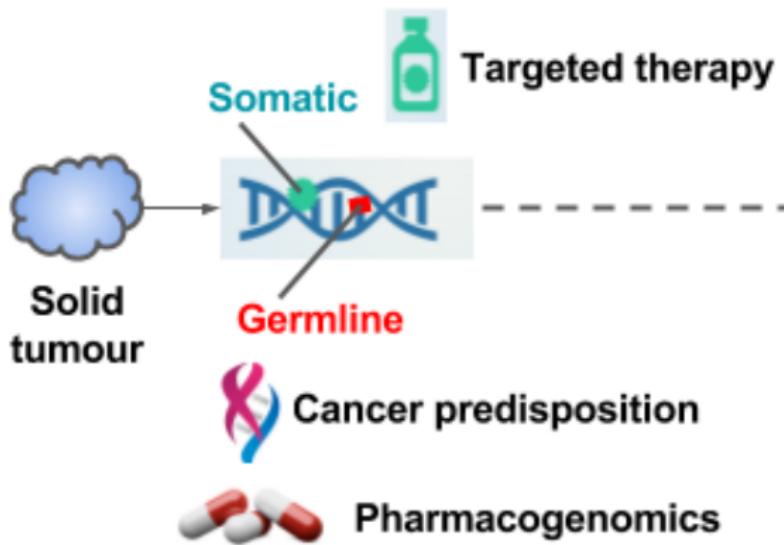
- ① Targeted next-generation sequencing panel for solid tumours

BC Cancer Agency's OncoPanel



- ① Targeted next-generation sequencing panel for solid tumours
- ② First gene panel to be available province wide and as part of standard of care

The tumour genome contains germline information



Clinical implications of germline variants

Cancer Predisposition

- Preventative measures
- Sibling testing



Pharmacogenomics

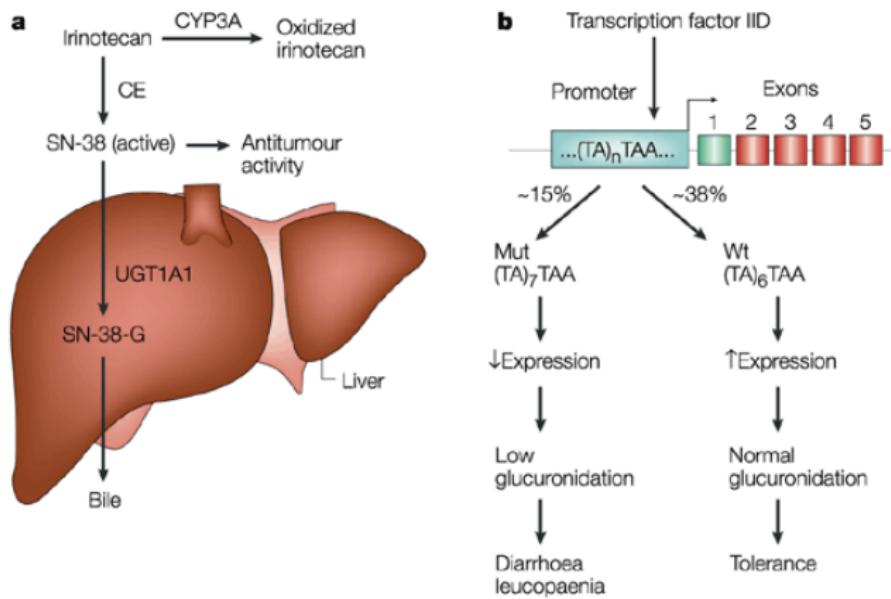
- Treatment tolerance
- Adverse drug reaction (ADR)

Patients with same diagnosis



- Full response
- No response
- Partial response
- ADR

Example: UGT1A1 promoter polymorphism leads to irinotecan toxicity



Nature Reviews | Cancer

¹Relling M.V & T. Dervieux, 2001, Nature Reviews Cancer 1, 99-108

Tumour-only sequencing is common in clinical laboratories

Clinical Laboratory



Matched Normal



- Blood
- Normal tissue



Tumours



Reason: Minimize cost and turnaround time

Overview

① Background

② Research Question

③ Project Aims

④ Aim 1

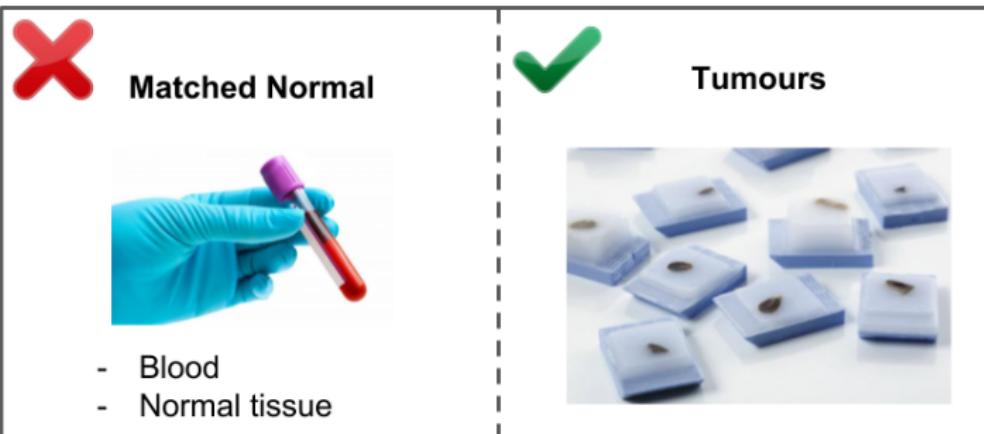
⑤ Aim 2

⑥ Aim 3

⑦ Conclusions

Research Question

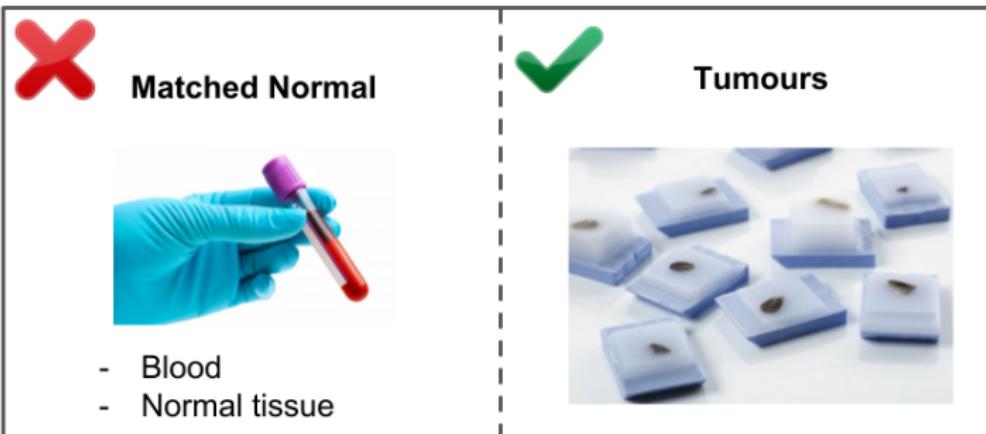
Clinical Laboratory



Reason: Minimize cost and turnaround time

Research Question

Clinical Laboratory



Reason: Minimize cost and turnaround time

Can we use tumour sequencing to screen for germline variants?

Fix the Fixation: Technical challenges in clinical genomics

Clinical Laboratory



Matched Normal



- Blood
- Normal tissue

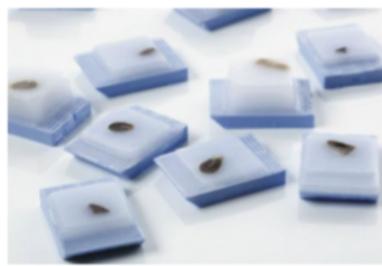


**Formalin-Fixed
Paraffin-Embedded
(FFPE) Tumours**



Fix the Fixation: Technical challenges in genetic testing

Clinical Laboratory

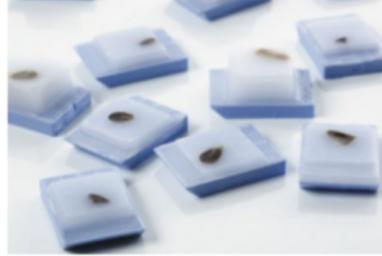
 Matched Normal  - Blood - Normal tissue	 Formalin-Fixed Paraffin-Embedded (FFPE) Tumours 
---	---

Formalin-induced artifacts:

- ① DNA fragmentation

Fix the Fixation: Technical challenges in genetic testing

Clinical Laboratory

 Matched Normal  - Blood - Normal tissue	 Formalin-Fixed Paraffin-Embedded (FFPE) Tumours 
---	---

Formalin-induced artifacts:

- ① DNA fragmentation
- ② Sequence artifacts (e.g. C>T/G>A transitions)

Recap

- ① The tumour genome contains germline information.

Recap

- ① The tumour genome contains germline information.
- ② Germline variants have clinical implications for patients and their families.

Recap

- ① The tumour genome contains germline information.
- ② Germline variants have clinical implications for patients and their families.
- ③ Matched normal samples are rarely processed by clinical laboratories due to additional **cost** and **turnaround time**.

Recap

- ① The tumour genome contains germline information.
- ② Germline variants have clinical implications for patients and their families.
- ③ Matched normal samples are rarely processed by clinical laboratories due to additional **cost** and **turnaround time**.

Research question: **Can we use tumour sequencing to screen for germline variants?**

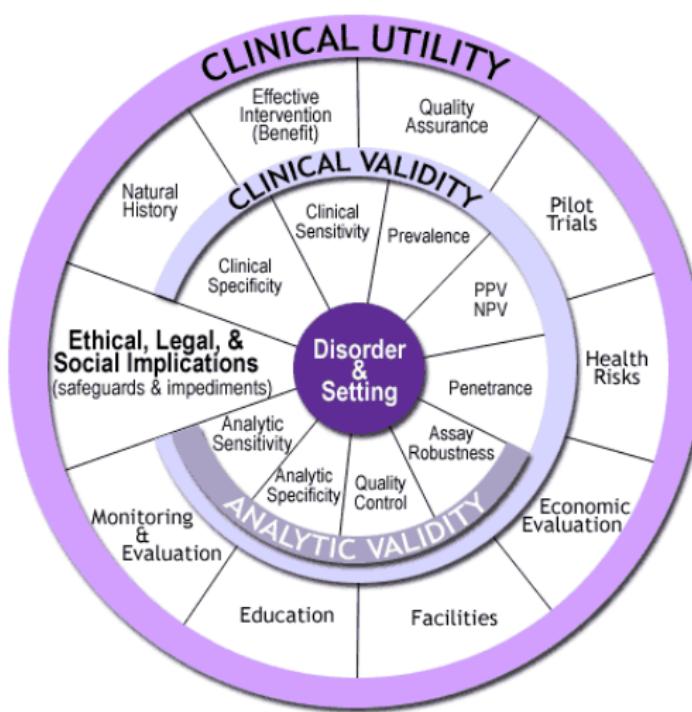
Recap

- ① The tumour genome contains germline information.
- ② Germline variants have clinical implications for patients and their families.
- ③ Matched normal samples are rarely processed by clinical laboratories due to additional **cost** and **turnaround time**.

Research question: **Can we use tumour sequencing to screen for germline variants?**

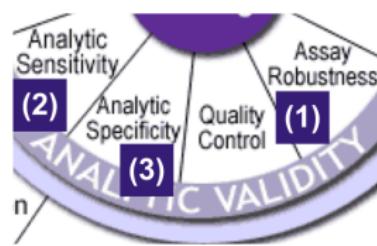
- ④ Tumour specimens are commonly FFPE, which causes to DNA damages.

Process for evaluation of genetic tests



Analytic Validation

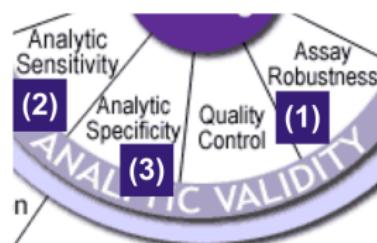
Can we use tumour sequencing to screen for germline variants?



- ① Do sequencing results differ between FFPE specimens and blood (gold standard for germline testing)?

Analytic Validation

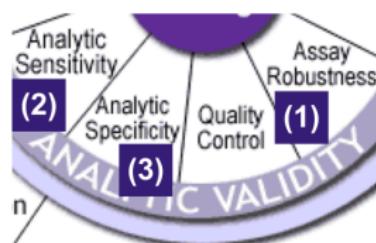
Can we use tumour sequencing to screen for germline variants?



- ① Do sequencing results differ between FFPE specimens and blood (gold standard for germline testing)?
- ② What is the true positive rate of detecting germline variants in FFPE tumours?

Analytic Validation

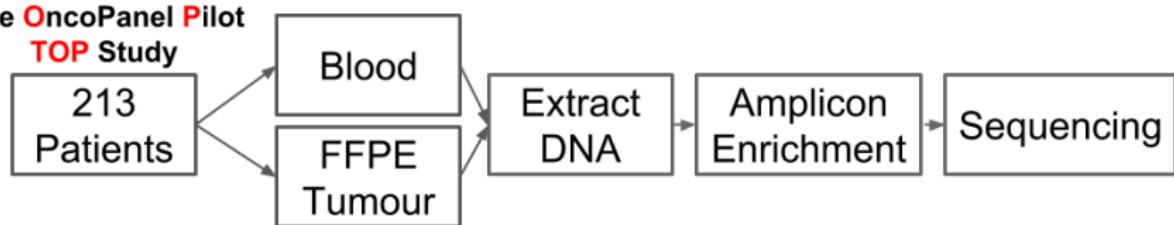
Can we use tumour sequencing to screen for germline variants?



- ① Do sequencing results differ between FFPE specimens and blood (gold standard for germline testing)?
- ② What is the true positive rate of detecting germline variants in FFPE tumours?
- ③ What is the percentage of true germline variants being referred to downstream germline testing (precision)?

Study Design

The OncoPanel Pilot
TOP Study



Tumour 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

Cancer-related genes in the OncoPanel

Gene	Protein
AKT1	Protein kinase B
ALK	Anaplastic lymphoma receptor tyrosine kinase
BRAF	Serine/threonine-protein kinase B-Raf
EGFR	Epidermal growth factor receptor
HRAS	GTPase HRas
MAPK1	Mitogen-activated protein kinase 1
MAP2K1	Mitogen-activated protein kinase kinase 1
MTOR	Serine/threonine-protein kinase mTOR
NRAS	Neuroblastoma RAS viral oncogene homolog
PDGFRA	Platelet-derived growth factor receptor alpha
PIK3CA	Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha
PTEN	Phosphatase and tensin homolog
STAT1	Signal transducer and activator of transcription 1
STAT3	Signal transducer and activator of transcription 3
TP53	Tumor protein P53

Pharmacogenomic genes in the OncoPanel

Gene	Protein	Chemotherapy
DPYD	Dihydropyrimidine dehydrogenase	5-FU
GSTM1	Glutathione S-transferase pi 1	Oxaliplatin
MTHFR	Methylenetetrahydrofolate reductase	5-FU
TYMP	Thymidine phosphorylase	5-FU
TYMS	Thymidylate synthetase	5-FU
UGT1A1	Uridine diphosphate (UDP)-glucuronosyl transferase 1A1	Irinotecan

Overview

① Background

② Research Question

③ Project Aims

④ Aim 1

⑤ Aim 2

⑥ Aim 3

⑦ Conclusions

Project Aims

Can we use tumour sequencing to screen for germline variants?

- ① **Assay QC:** Compare efficiency in amplicon enrichment and sequencing results between blood and FFPE specimens

Project Aims

Can we use tumour sequencing to screen for germline variants?

- ① **Assay QC:** Compare efficiency in amplicon enrichment and sequencing results between blood and FFPE specimens
- ② **Sensitivity:** Determine the true positive rate for detection of germline variants in FFPE tumours

Project Aims

Can we use tumour sequencing to screen for germline variants?

- ① **Assay QC:** Compare efficiency in amplicon enrichment and sequencing results between blood and FFPE specimens
- ② **Sensitivity:** Determine the true positive rate for detection of germline variants in FFPE tumours
- ③ **Precision:** Determine the percentage of true germline variants referred for downstream germline testing

Overview

1 Background

2 Research Question

3 Project Aims

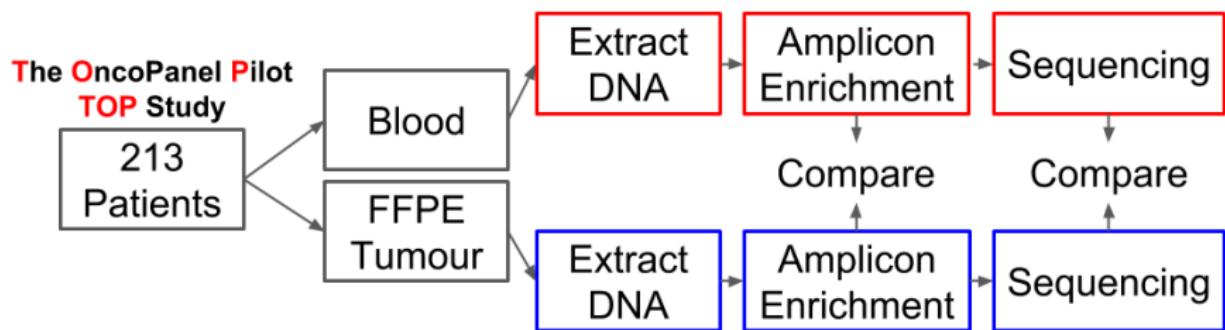
4 Aim 1

5 Aim 2

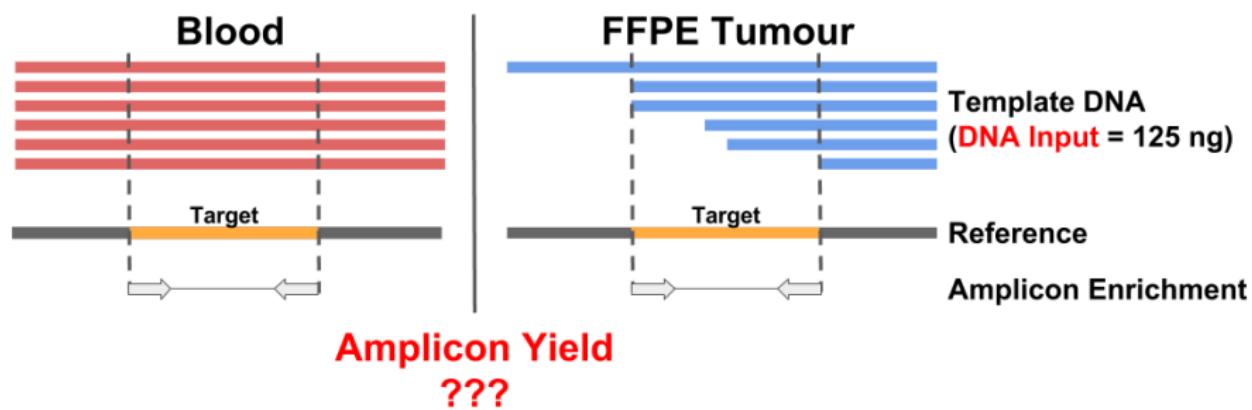
6 Aim 3

7 Conclusions

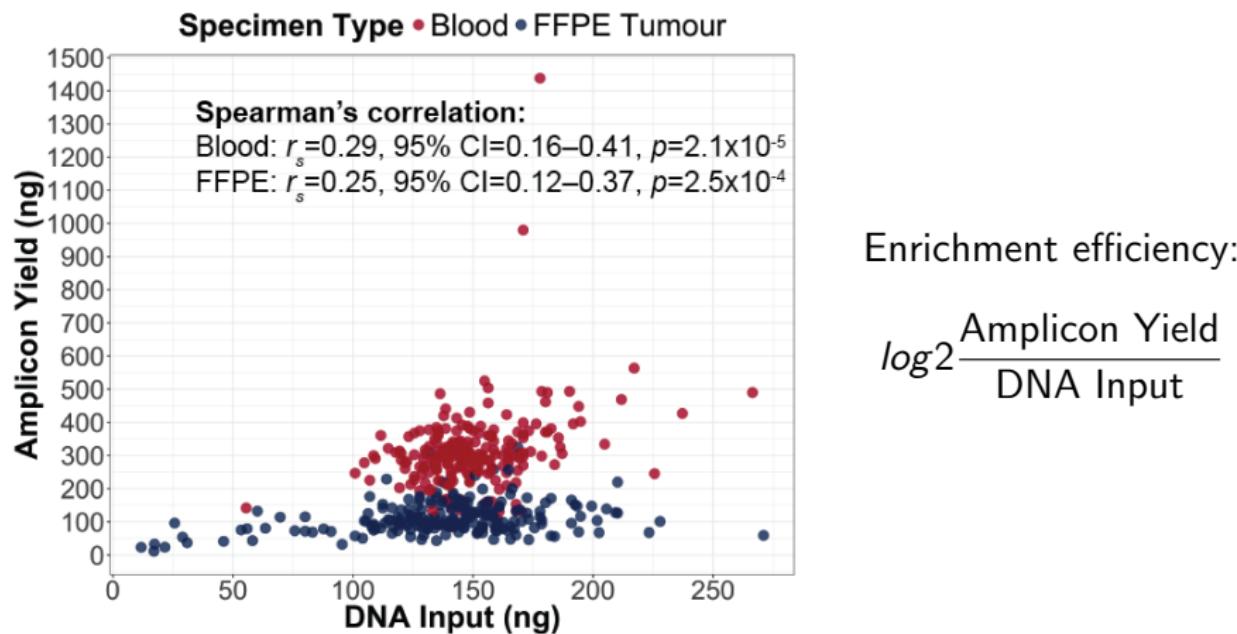
Aim 1: Compare efficiency in amplicon enrichment and sequencing results between blood and FFPE specimens



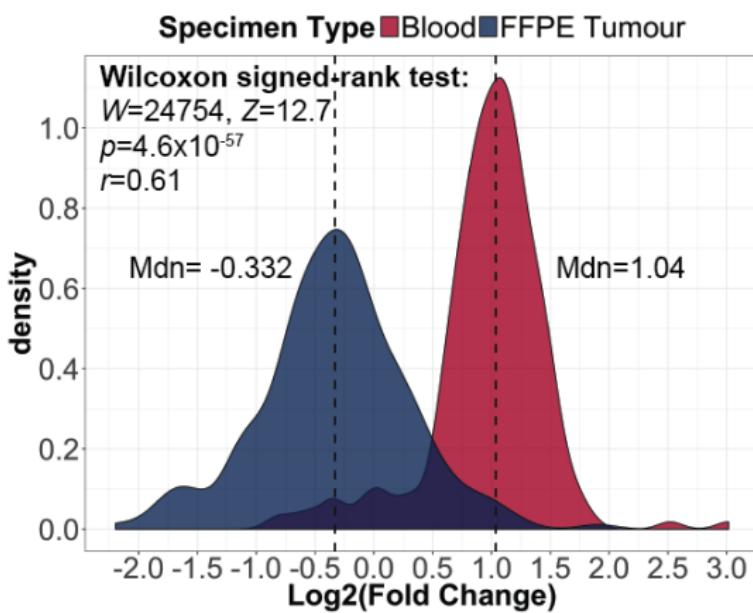
Formalin fixation causes DNA fragmentation



Correlation between DNA input and amplicon yield



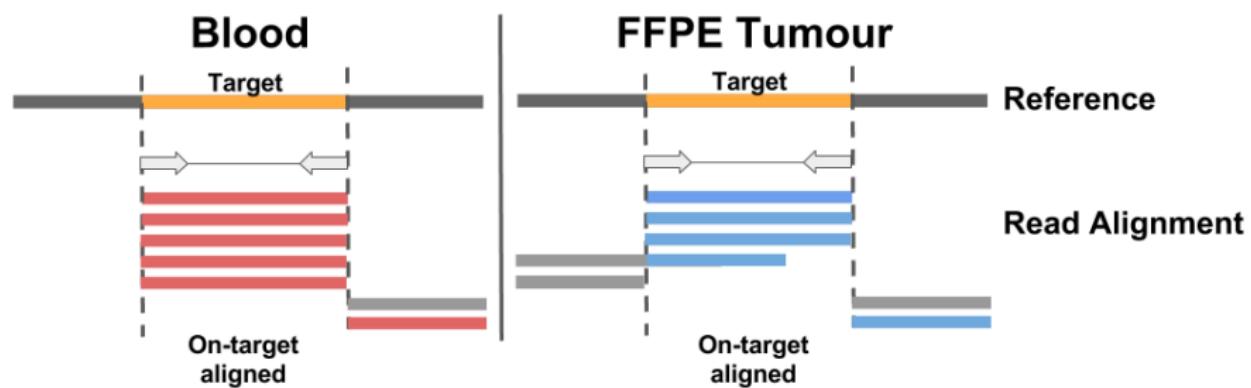
Reduced efficiency in amplicon enrichment is observed in FFPE specimens



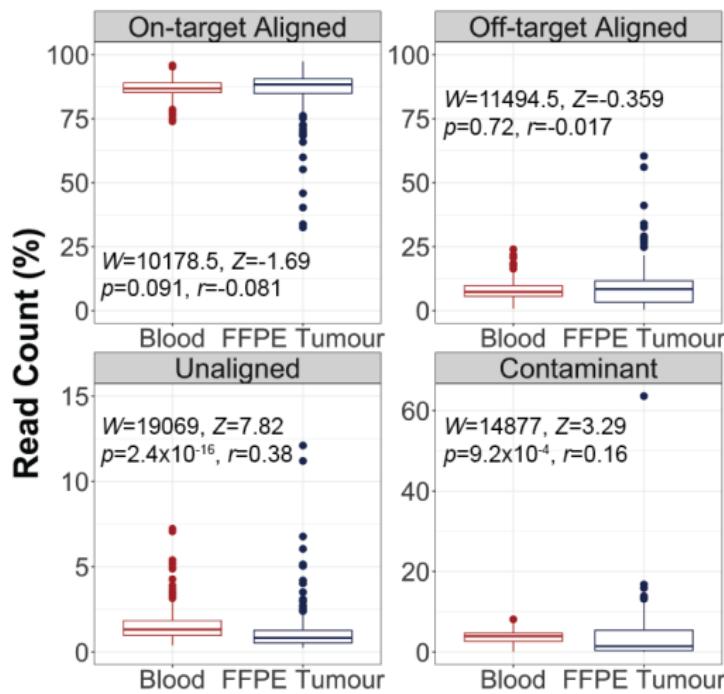
Enrichment efficiency:

$$\log_2 \frac{\text{Amplicon Yield}}{\text{DNA Input}}$$

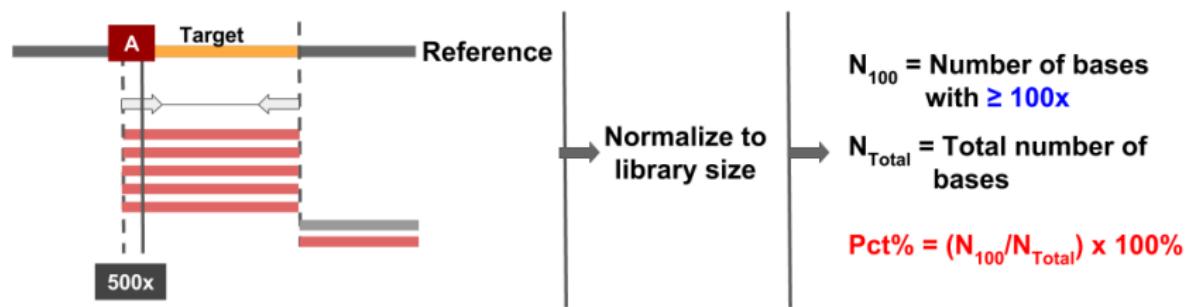
Read Alignments



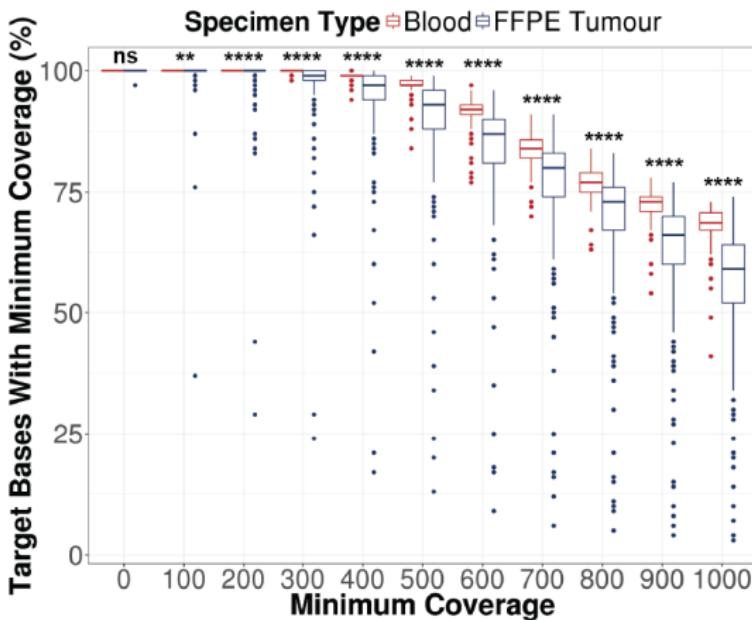
Percentage of on-target aligned reads is comparable between specimen types



Per base coverage statistics

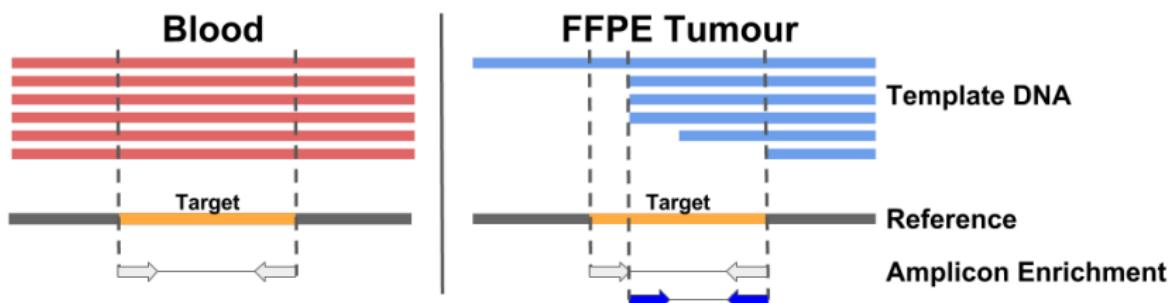


Percentage of target bases is significantly different at all coverage thresholds



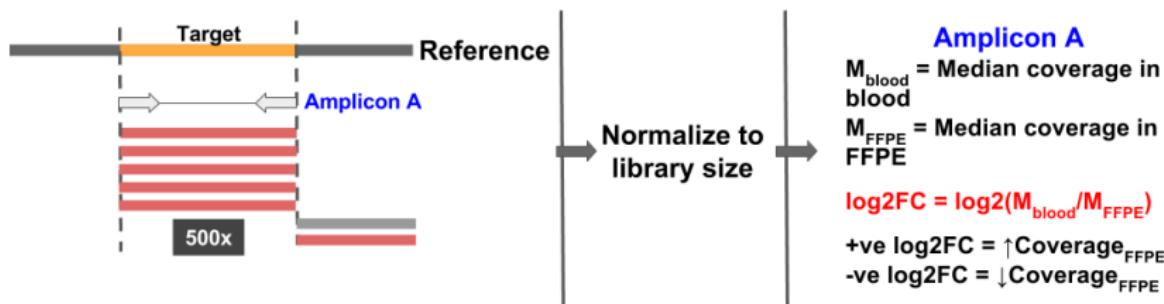
Wilcoxon signed-rank test, **** $p < 0.0001$, *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, ns = not significant

OncoPanel consists of 416 amplicons



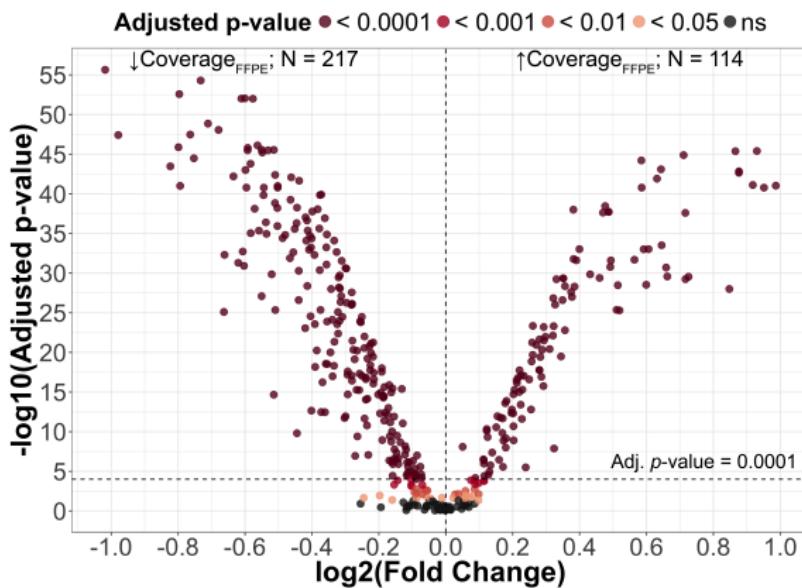
Shorter amplicons might yield greater coverage depth in FFPE specimens due to fragmentation damages in template DNA.

Analysis of Amplicon Coverage Depth



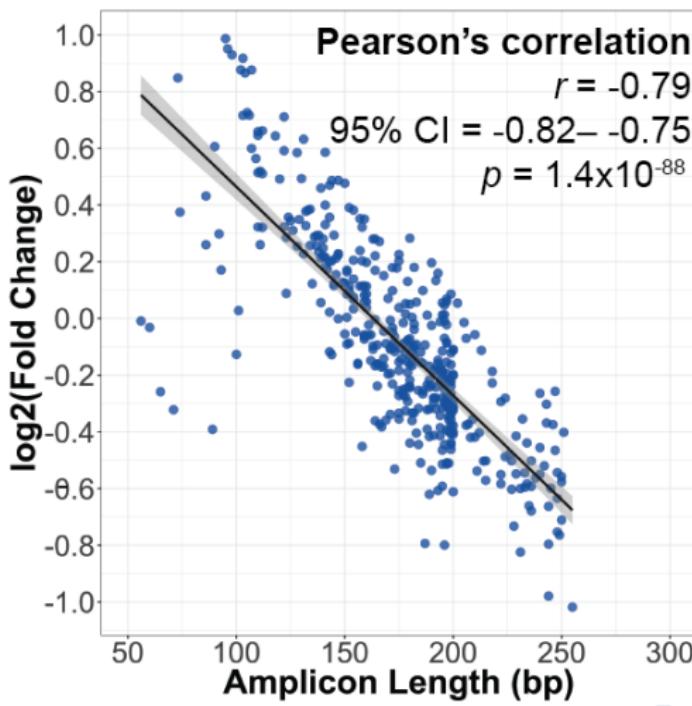
Comparison of amplicon coverage was performed with the Wilcoxon signed-rank test.

There are more amplicons with lower coverage depth in FFPE specimens relative to blood specimens

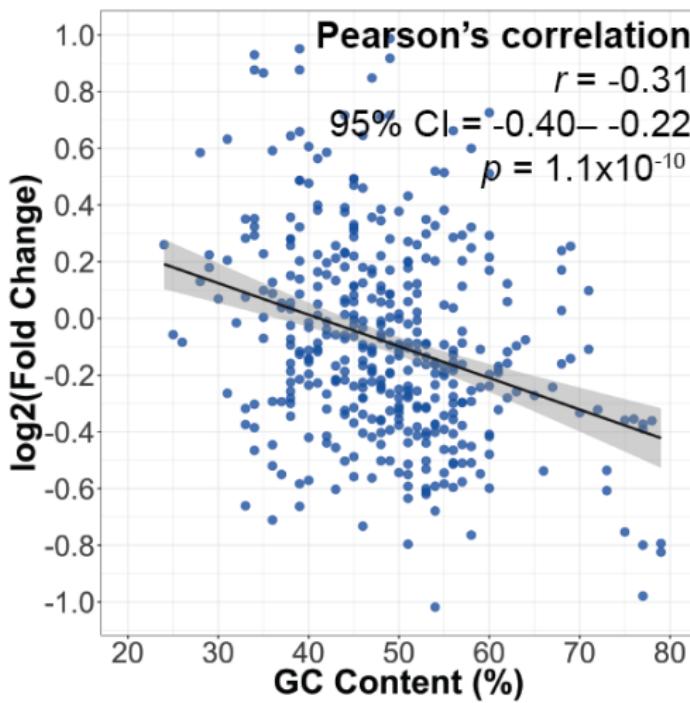


Adj. p-value = Wilcoxon signed-rank test with Benjamini Hochberg correction

Decreased amplicon coverage in FFPE specimens is correlated with increased amplicon length



Decreased amplicon coverage in FFPE specimens is correlated with increased amplicon GC content



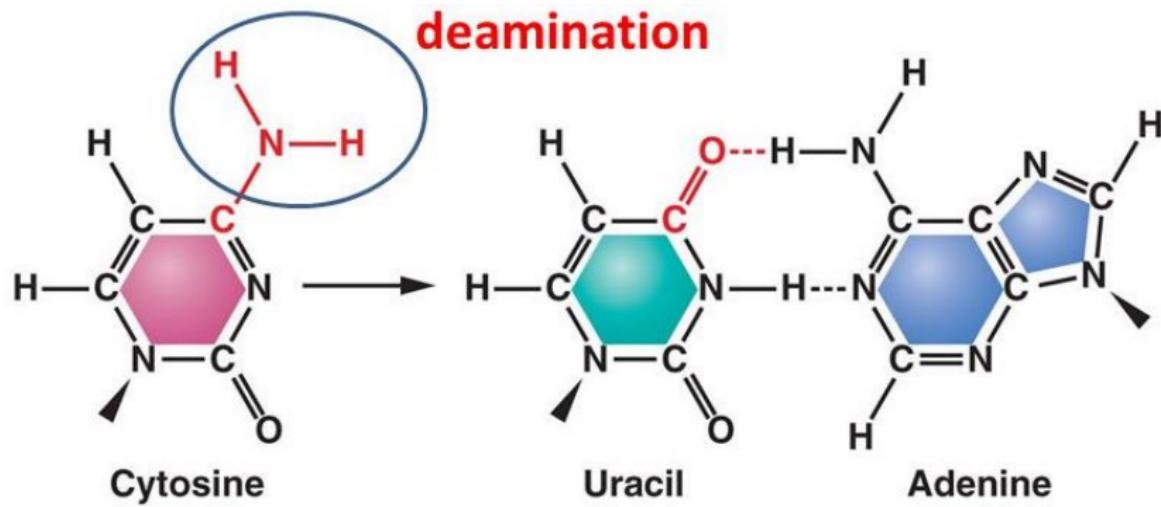
Reduced amplicon coverage in FFPE specimens is more pronounced for longer amplicons

Multiple regression of amplicon length and GC content as predictors for log2FC between amplicon coverage in FFPE specimens and blood.

Variable	Unstandardized	S.E.	Standardized	p-value
Length (bp)	-7.24×10^{-3}	2.54×10^{-4}	-7.75×10^{-1}	2.47×10^{-99}
GC Content (%)	-9.92×10^{-3}	9.77×10^{-4}	-2.77×10^{-1}	8.70×10^{-22}
Intercept = 1.66, Adjusted R ² = 0.695 $F(2, 411) = 471$, p-value = 4.65×10^{-107}				

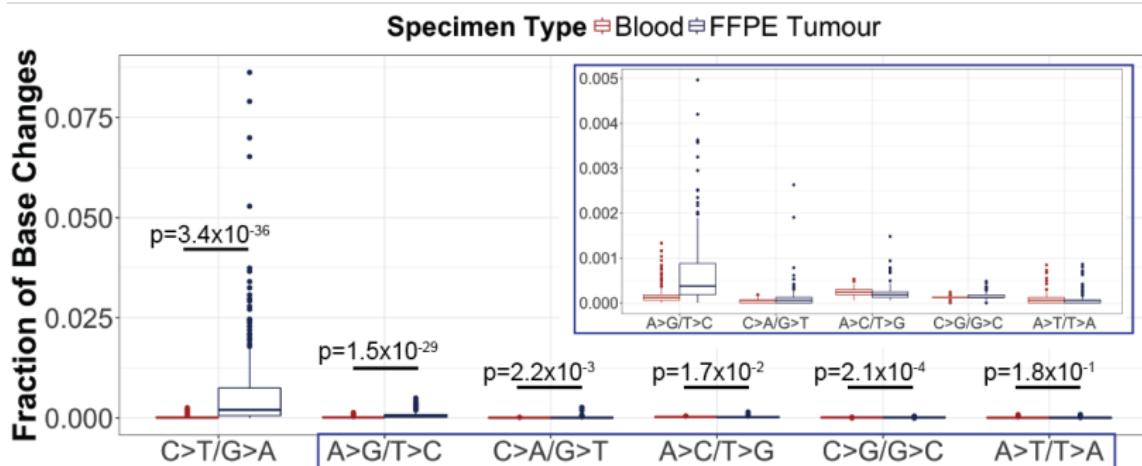
A change in 1 standard deviation of amplicon length has >2x the impact on log2FC than a 1 standard deviation change in amplicon GC content.

Formalin fixation induces deamination of cytosine bases

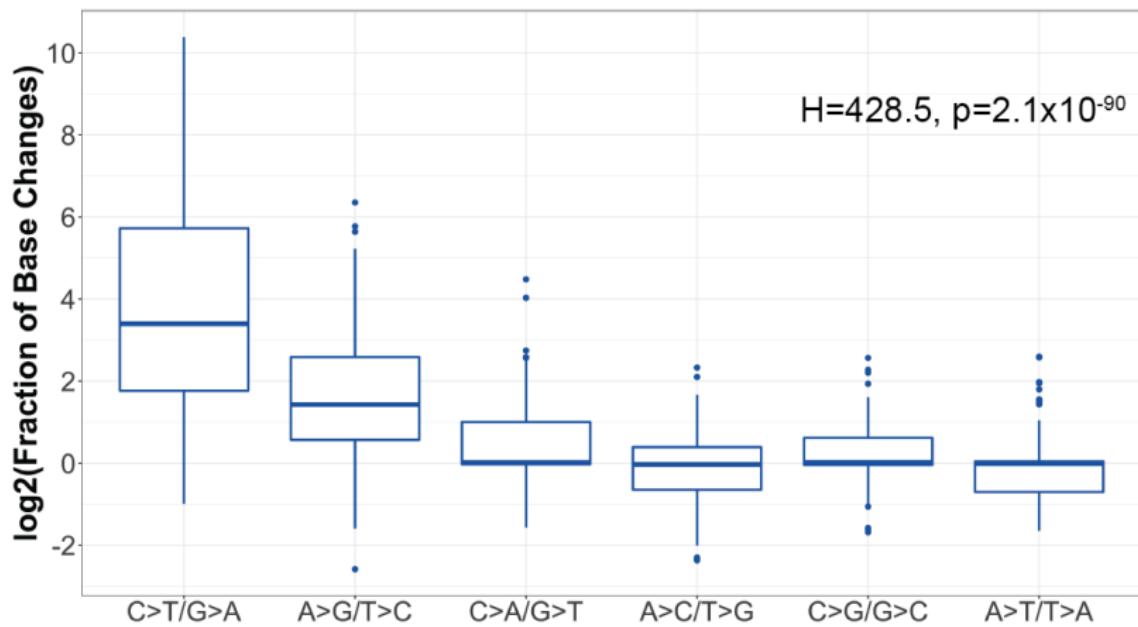


¹Klug and Cummings, 1997

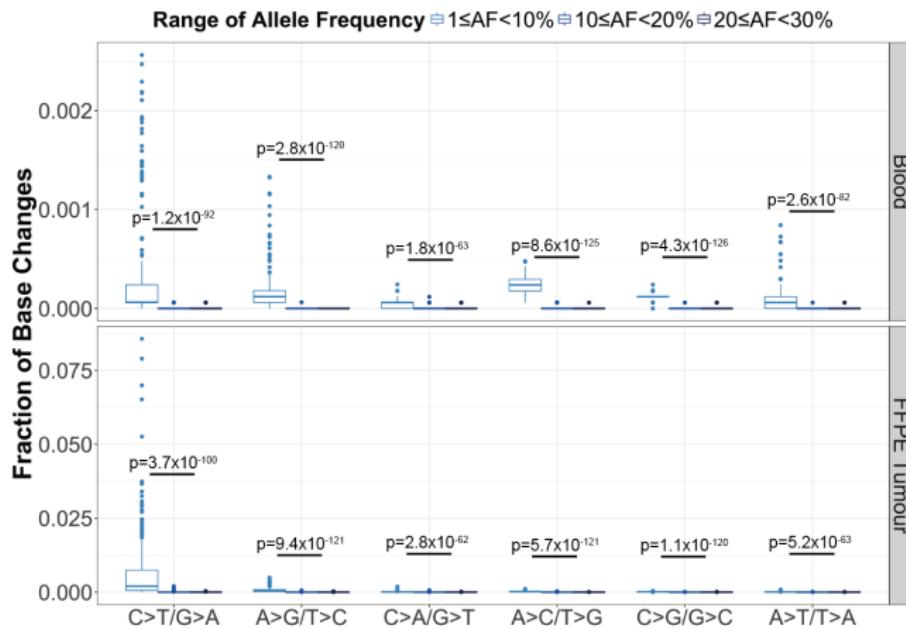
Deamination effects lead to increased C>T/G>A transitions in FFPE specimens (Wilcoxon signed-rank test)



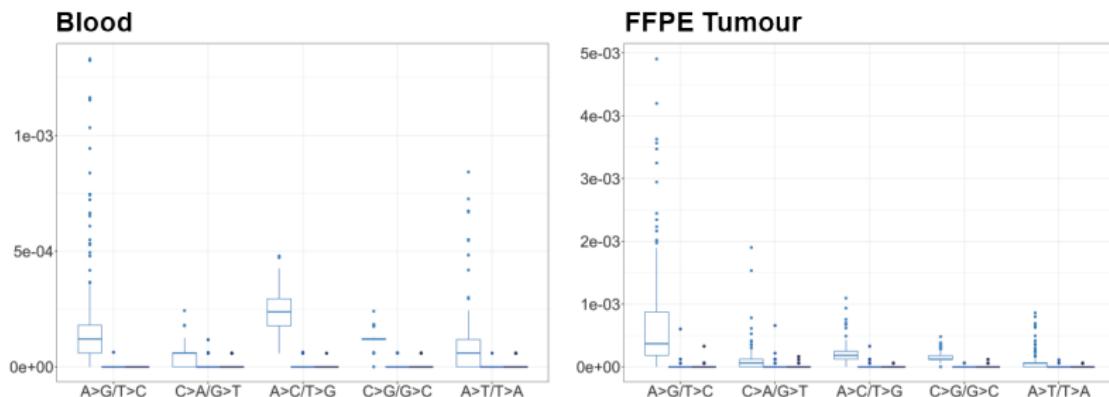
Deamination effects lead to increased C>T/G>A transitions in FFPE specimens (Wilcoxon signed-rank test, fold change)



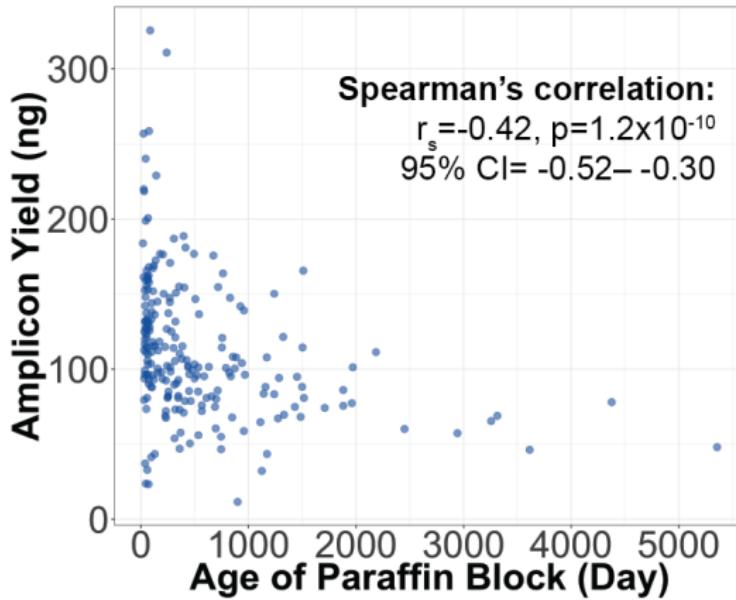
Deamination effects lead to increased C>T/G>A transitions in FFPE specimens at low allele frequency (Wilcoxon signed-rank test)



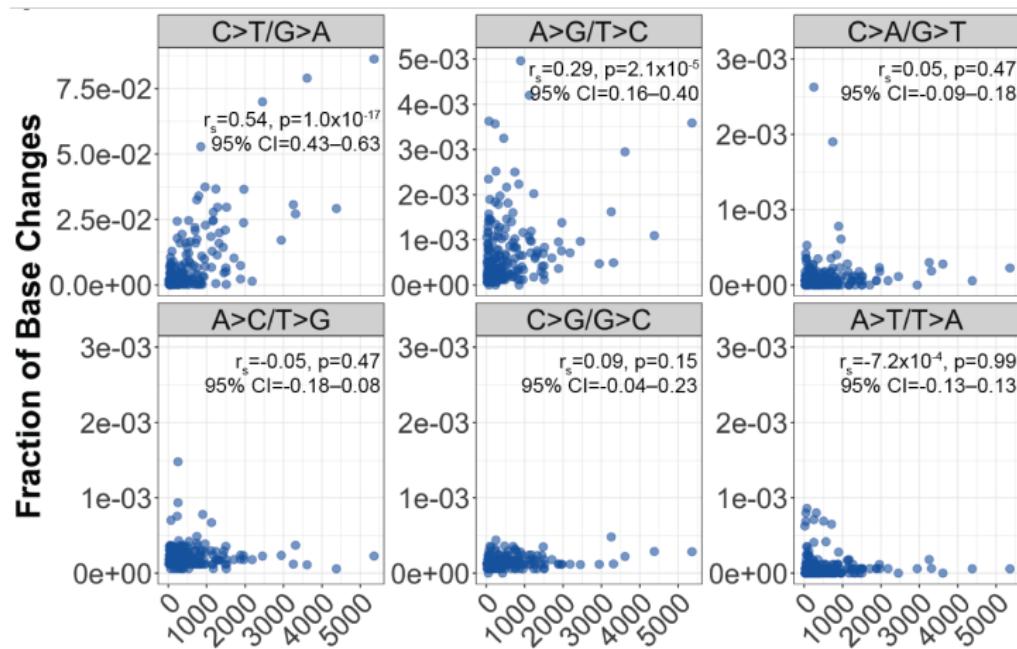
Deamination effects lead to increased C>T/G>A transitions in FFPE specimens at low allele frequency
(Wilcoxon signed-rank test, enlarged figure)



Increased age of paraffin block results in reduced amplicon yield



Increased age of paraffin block results in elevated events of C>T/G>A sequence artifacts (Spearman's correlation)



Determination of correlation between pre-sequencing variables and sequencing results using Spearman's correlation. Top values represent Spearman's *rho* and 95% confidence interval in brackets, whereas bottom values represent *p*-value. Asterisk(*) indicates significance level of *p*-value < 0.05.

Variable	Amplicon Yield (ng)	Age of Paraffin Block (Day)	Fraction of C>T/G>A	Average Per Base Normalized Coverage
Age of Paraffin Block (Day)	-0.42 (-0.52– -0.30) $5.2 \times 10^{-7}*$			
Fraction of C>T/G>A	-0.72 (-0.77– -0.65) $1.9 \times 10^{-11}*$	0.54 (0.61–0.75) $6.3 \times 10^{-35}*$		
Average Per Base Normalized Coverage	0.69 (0.61–0.75) $8.5 \times 10^{-20}*$	-0.47 (-0.57– -0.36) $4.7 \times 10^{-7}*$	-0.80 (-0.84– -0.75) $7.5 \times 10^{-17}*$	
On-target Aligned Reads (%)	0.58 (0.48–0.66) $2.1 \times 10^{-13}*$	-0.35 (-0.46– -0.23) $8.2 \times 10^{-3}*$	-0.57 (-0.65– -0.47) $4.2 \times 10^{-8}*$	0.73 (0.66–0.79) $3.1 \times 10^{-58}*$

Summary for Aim 1

- ① Efficiency in amplicon enrichment is lower in FFPE compared to blood.

Summary for Aim 1

- ① Efficiency in amplicon enrichment is lower in FFPE compared to blood.
- ② Percentage of on-target aligned reads (%) is comparable between blood and FFPE specimens.

Summary for Aim 1

- ① Efficiency in amplicon enrichment is lower in FFPE compared to blood.
- ② Percentage of on-target aligned reads (%) is comparable between blood and FFPE specimens.
- ③ There is discrepancy in coverage depth between blood and FFPE specimens.

Summary for Aim 1

- ① Efficiency in amplicon enrichment is lower in FFPE compared to blood.
- ② Percentage of on-target aligned reads (%) is comparable between blood and FFPE specimens.
- ③ There is discrepancy in coverage depth between blood and FFPE specimens.
- ④ Shorter amplicons achieve greater coverage in FFPE specimens compared to longer amplicons.

Summary for Aim 1

- ① Efficiency in amplicon enrichment is lower in FFPE compared to blood.
- ② Percentage of on-target aligned reads (%) is comparable between blood and FFPE specimens.
- ③ There is discrepancy in coverage depth between blood and FFPE specimens.
- ④ Shorter amplicons achieve greater coverage in FFPE specimens compared to longer amplicons.
- ⑤ Increased C>T/G>A transitions are observed in FFPE specimens, and this increase is correlated with increased age of paraffin blocks.

Overview

1 Background

2 Research Question

3 Project Aims

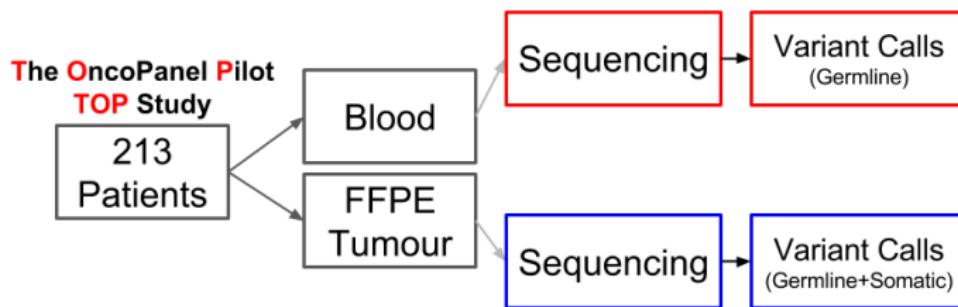
4 Aim 1

5 Aim 2

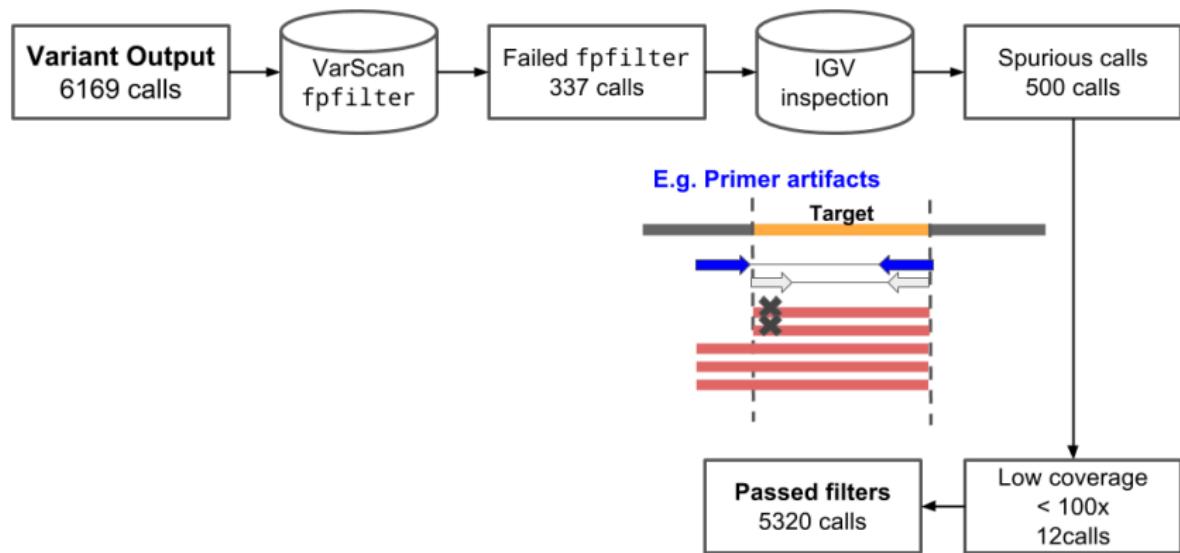
6 Aim 3

7 Conclusions

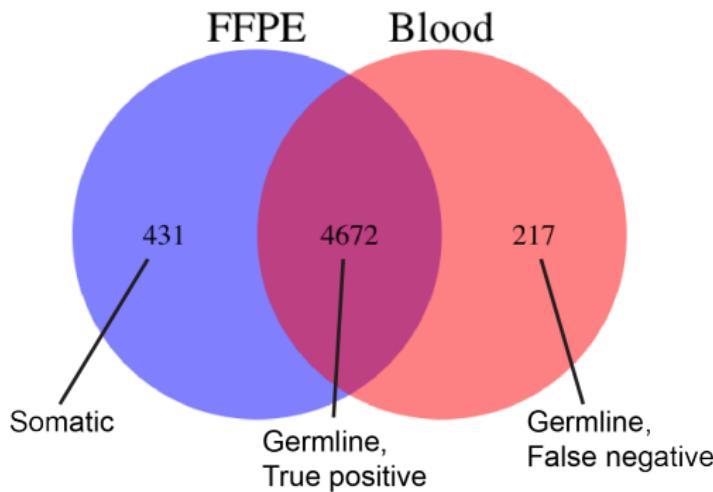
Aim 2: Determine the true positive rate for detection of germline variants in FFPE tumours (sensitivity)



Variant calling pipeline



Germline variants are highly concordant between blood and FFPE specimens



$$\text{True positive rate} = 4672 / (217 + 4672) = 0.956$$

Overview

① Background

② Research Question

③ Project Aims

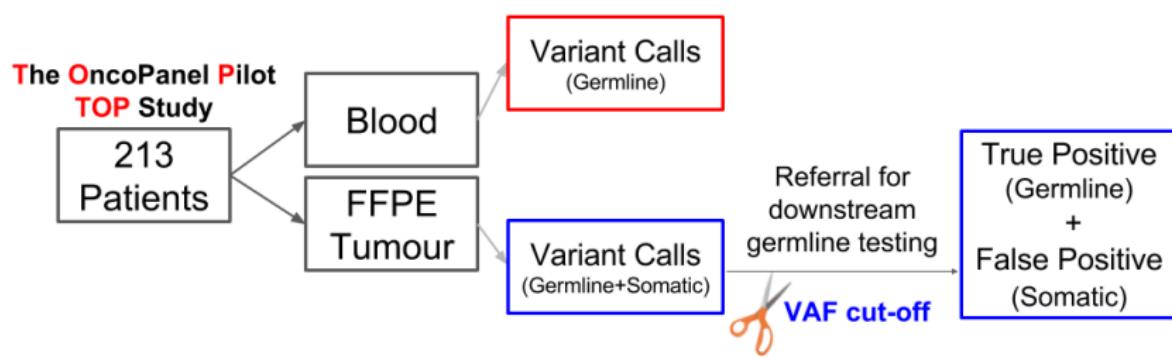
④ Aim 1

⑤ Aim 2

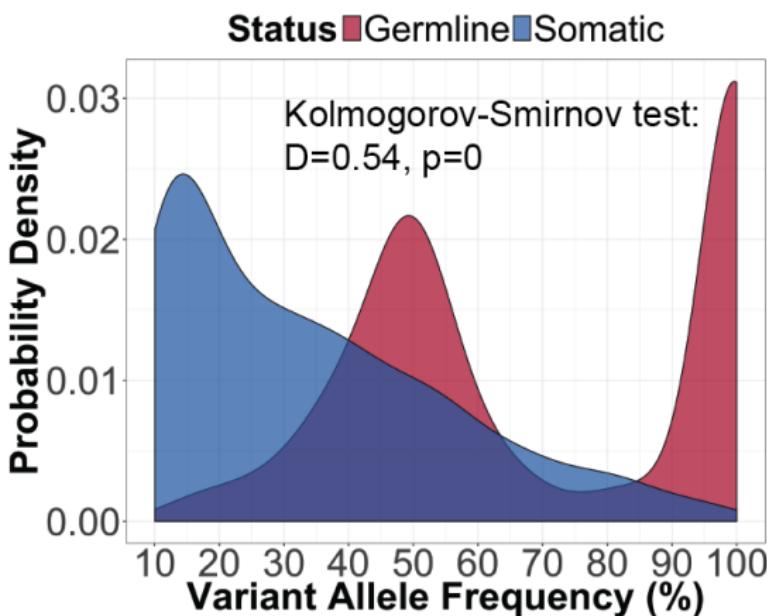
⑥ Aim 3

⑦ Conclusions

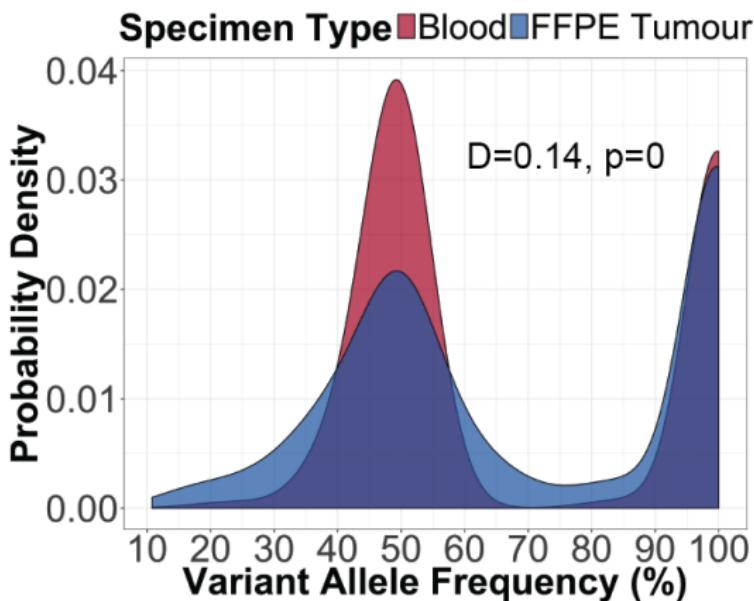
Aim 3: Determine the percentage of true germline variants referred for downstream germline testing (precision/positive predictive value)



VAF of germline and somatic variants in FFPE tumour



VAF of germline variants in blood and FFPE tumour

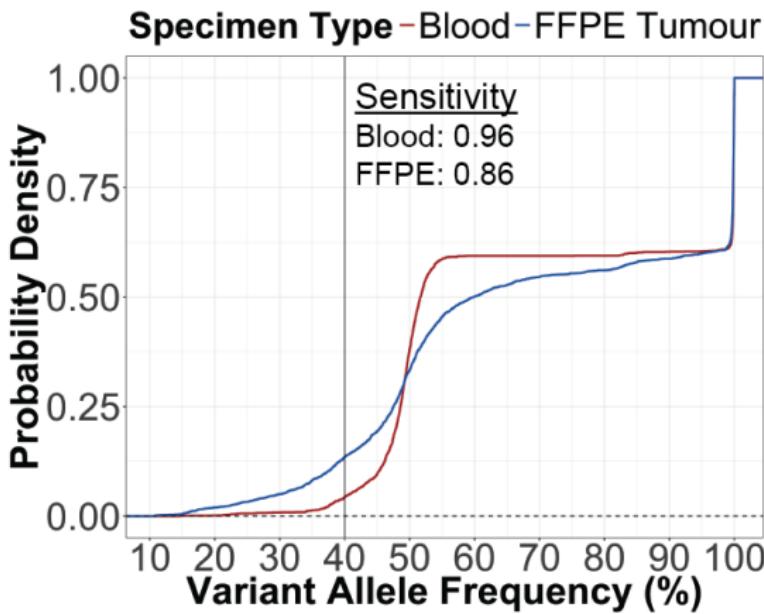


Reduced sensitivity in detection of germline variants at 40% VAF

VAF (%)	Blood				FFPE Tumour			
	FN*	TP**	Sensitivity	95% CI	FN*	TP**	Sensitivity	95% CI
10	0	2461	1.0	1.0–1.0	0	2428	1.0	1.0–1.0
15	2	2459	1.0	1.0–1.0	12	2416	1.0	0.99–1.0
20	3	2458	1.0	1.0–1.0	48	2380	0.98	0.97–0.99
25	15	2446	0.99	0.99–1.00	79	2349	0.97	0.96–0.97
30	20	2441	0.99	0.99–1.00	121	2307	0.95	0.94–0.96
35	33	2428	0.99	0.98–0.99	197	2231	0.92	0.91–0.93
40	107	2354	0.96	0.95–0.96	328	2100	0.86	0.85–0.88
45	234	2227	0.90	0.89–0.92	470	1958	0.81	0.79–0.82

*FN = False negative; **TP = True positive

Reduced sensitivity in detection of germline variants at 40% VAF

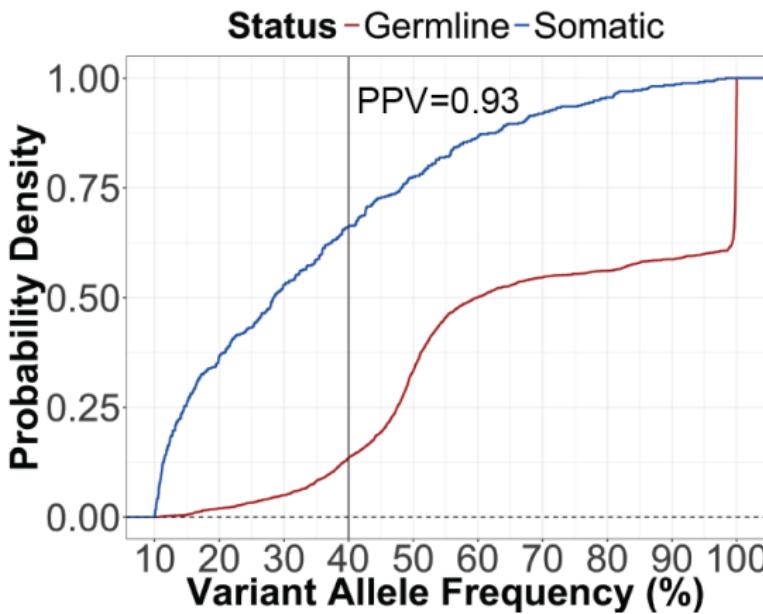


High positive predictive value can be achieved for referral of germline variants to downstream confirmatory testing

VAF (%)	FP*	TP**	Total Calls	Positive Predictive Value	95% CI
10	431	2428	2859	0.85	0.84–0.86
15	319	2416	2735	0.88	0.87–0.90
20	273	2380	2653	0.90	0.88–0.91
25	245	2349	2594	0.91	0.89–0.92
30	203	2307	2510	0.92	0.91–0.93
35	178	2231	2409	0.93	0.91–0.94
40	146	2100	2246	0.93	0.92–0.94
45	118	1958	2076	0.94	0.93–0.95

*FP = False positive; **TP = True positive

High positive predictive value can be achieved for referral of germline variants to downstream confirmatory testing



Overview

1 Background

2 Research Question

3 Project Aims

4 Aim 1

5 Aim 2

6 Aim 3

7 Conclusions

Conclusions

- ① Formalin-induced DNA damages are detectable in NGS data.

Conclusions

- ① Formalin-induced DNA damages are detectable in NGS data.
- ② Germline variants are highly concordant between blood and FFPE specimens ($\text{TPR} = 0.956$).

Conclusions

- ① Formalin-induced DNA damages are detectable in NGS data.
- ② Germline variants are highly concordant between blood and FFPE specimens (**TPR = 0.956**).
- ③ At 40% VAF threshold, sensitivity for detection of germline variants in FFPE tumour is **0.86** and the positive predictive value for referral to downstream confirmatory testing is **0.93**.

Acknowledgements

Karsan lab past and current members

Committee:

- Dr. Ryan Morin
- Dr. Martin Hirst

Centre of Clinical Genomics:

- Liz Starks
- Jillian Slind
- Hadrien Jouet

Acknowledgements

GSC sequencing team

Cancer Genetics Lab

Patients

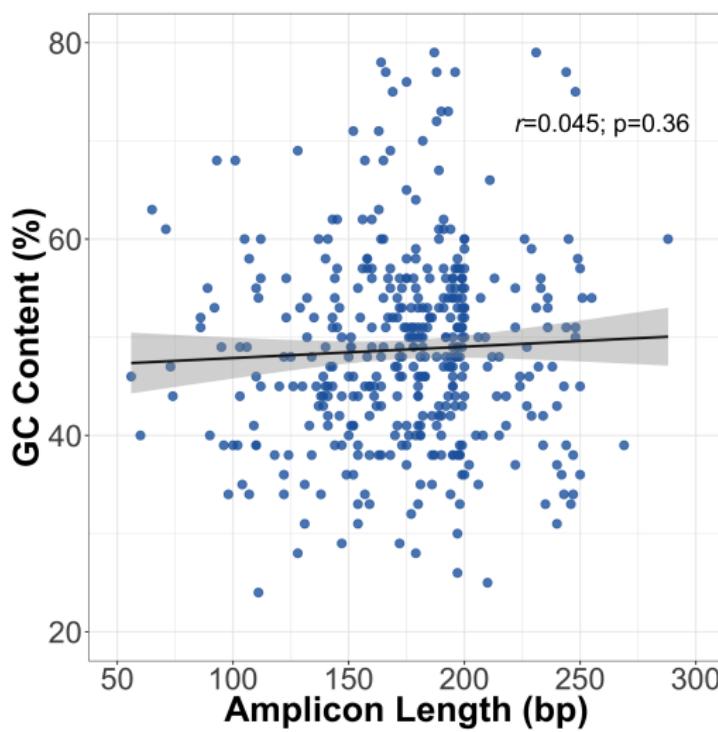
Acknowledgements



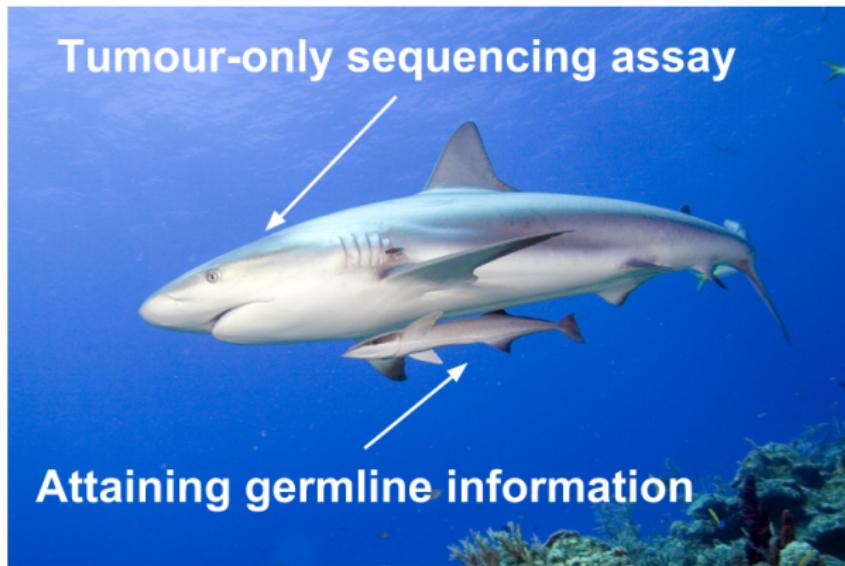
Percentage of target bases is significantly different at all coverage thresholds

Threshold	Blood		FFPE Tumour		$p (< 0.05^*)$	r
	Median (%)	Range (%)	Median (%)	Range (%)		
$\geq 0x$	100.0	100.0–100.0	100.0	97.0–100.0	1.0	0.068
$\geq 100x$	100.0	100.0–100.0	100.0	37.0–100.0	$2.3 \times 10^{-4}^*$	0.25
$\geq 200x$	100.0	100.0–100.0	100.0	29.0–100.0	$2.9 \times 10^{-11}^*$	0.41
$\geq 300x$	100.0	98.0–100.0	99.0	24.0–100.0	$4.1 \times 10^{-18}^*$	0.55
$\geq 400x$	99.0	94.0–100.0	97.0	17.0–100.0	$5.0 \times 10^{-28}^*$	0.68
$\geq 500x$	97.0	84.0–99.0	89.5	13.0–99.0	$2.1 \times 10^{-38}^*$	0.77
$\geq 600x$	92.0	77.0–97.0	87.0	9.0–96.0	$1.5 \times 10^{-32}^*$	0.72
$\geq 700x$	84.0	70.0–91.0	80.0	6.0–91.0	$5.7 \times 10^{-25}^*$	0.65
$\geq 800x$	77.0	63.0–84.0	73.0	5.0–83.0	$4.7 \times 10^{-27}^*$	0.67
$\geq 900x$	73.0	54.0–78.0	66.0	4.0–77.0	$4.6 \times 10^{-40}^*$	0.78
$\geq 1000x$	68.5	41.0–73.0	59.0	3.0–74.0	$3.6 \times 10^{-42}^*$	0.79

Amplicon length and GC content



Research Question



Tumour-only sequencing assay

Attaining germline information

Can we leverage tumour genomic testing to perform screening for clinically relevant germline variants?