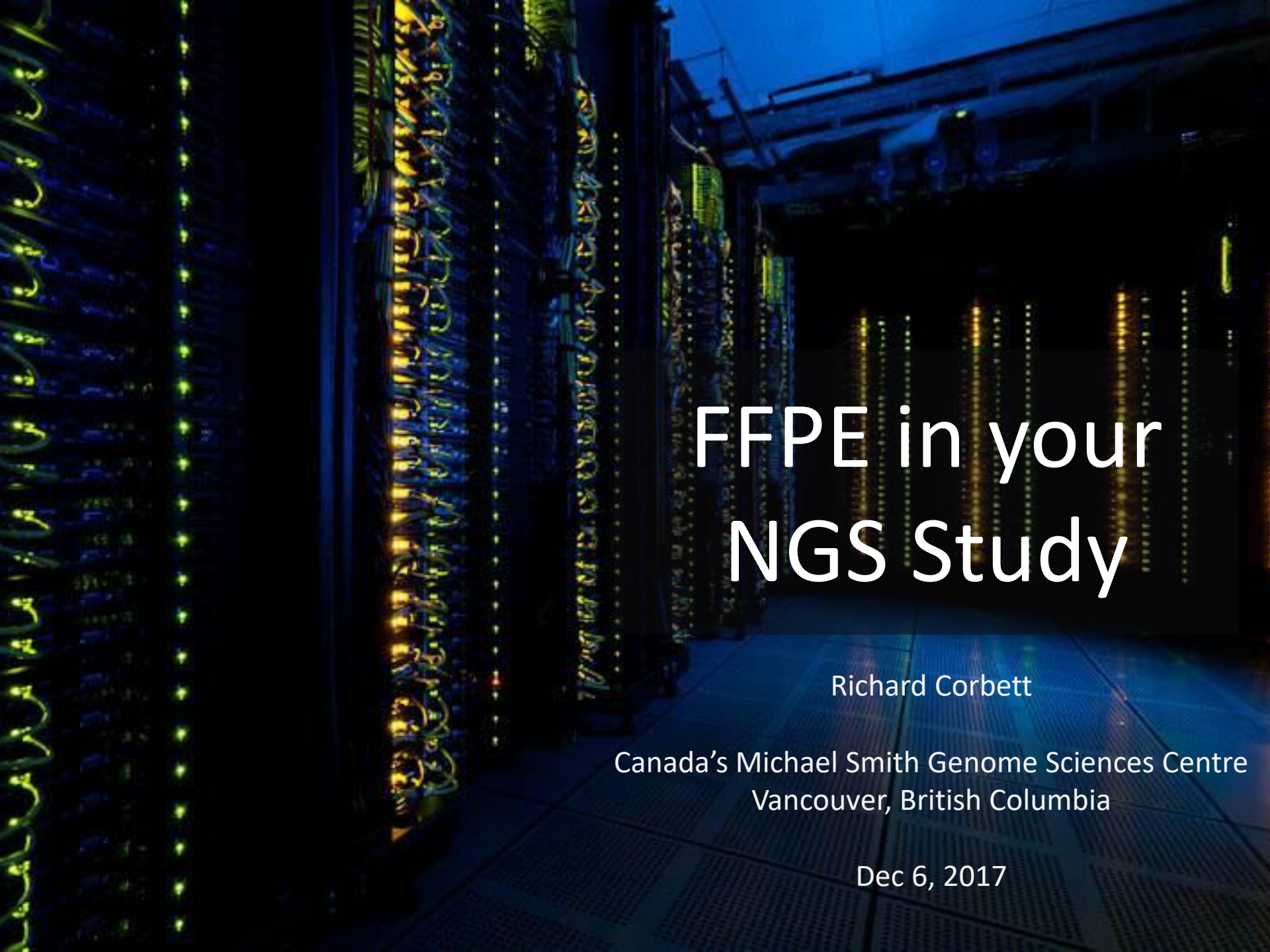
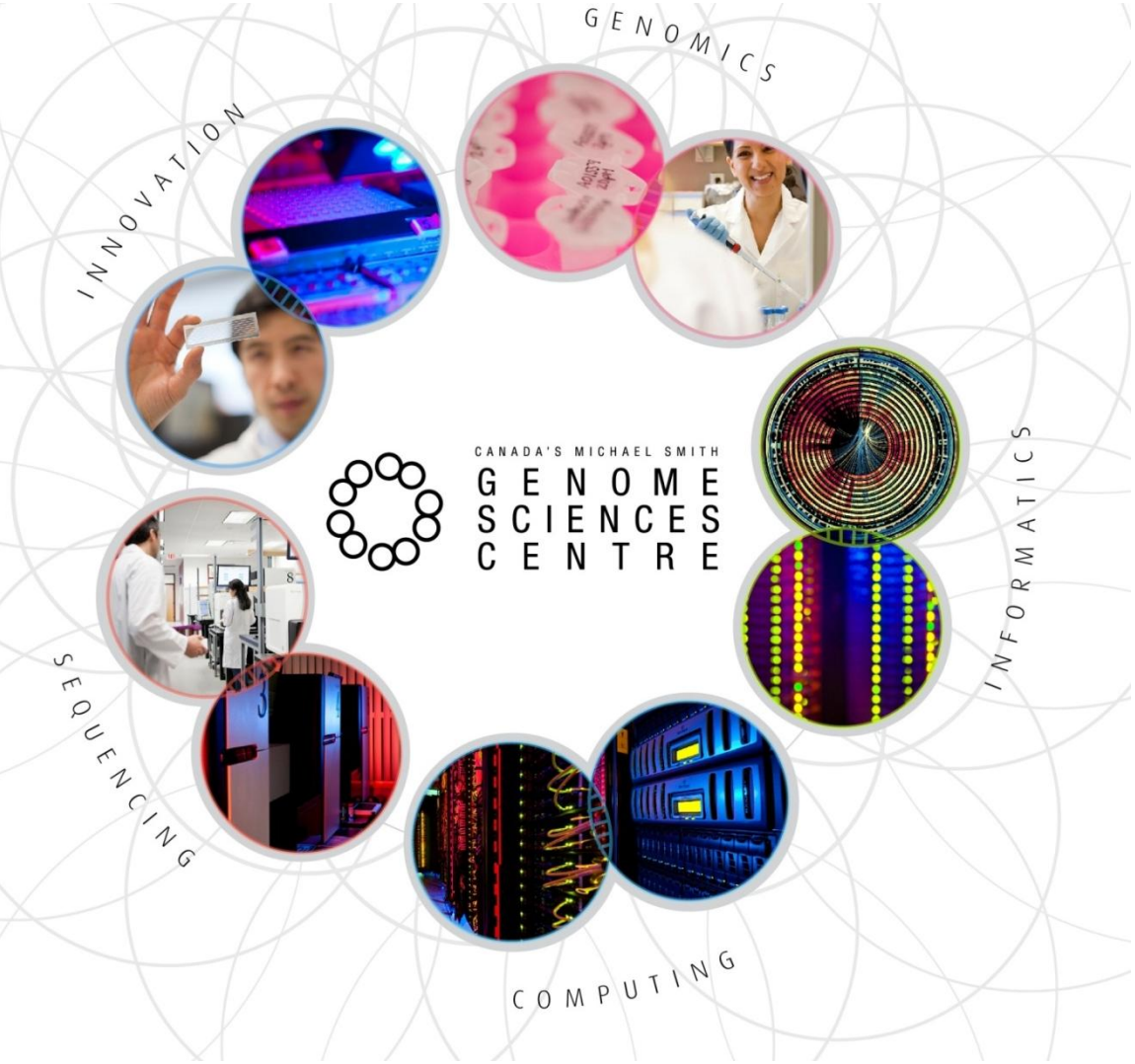


FFPE in your NGS Study

Richard Corbett

Canada's Michael Smith Genome Sciences Centre
Vancouver, British Columbia

Dec 6, 2017



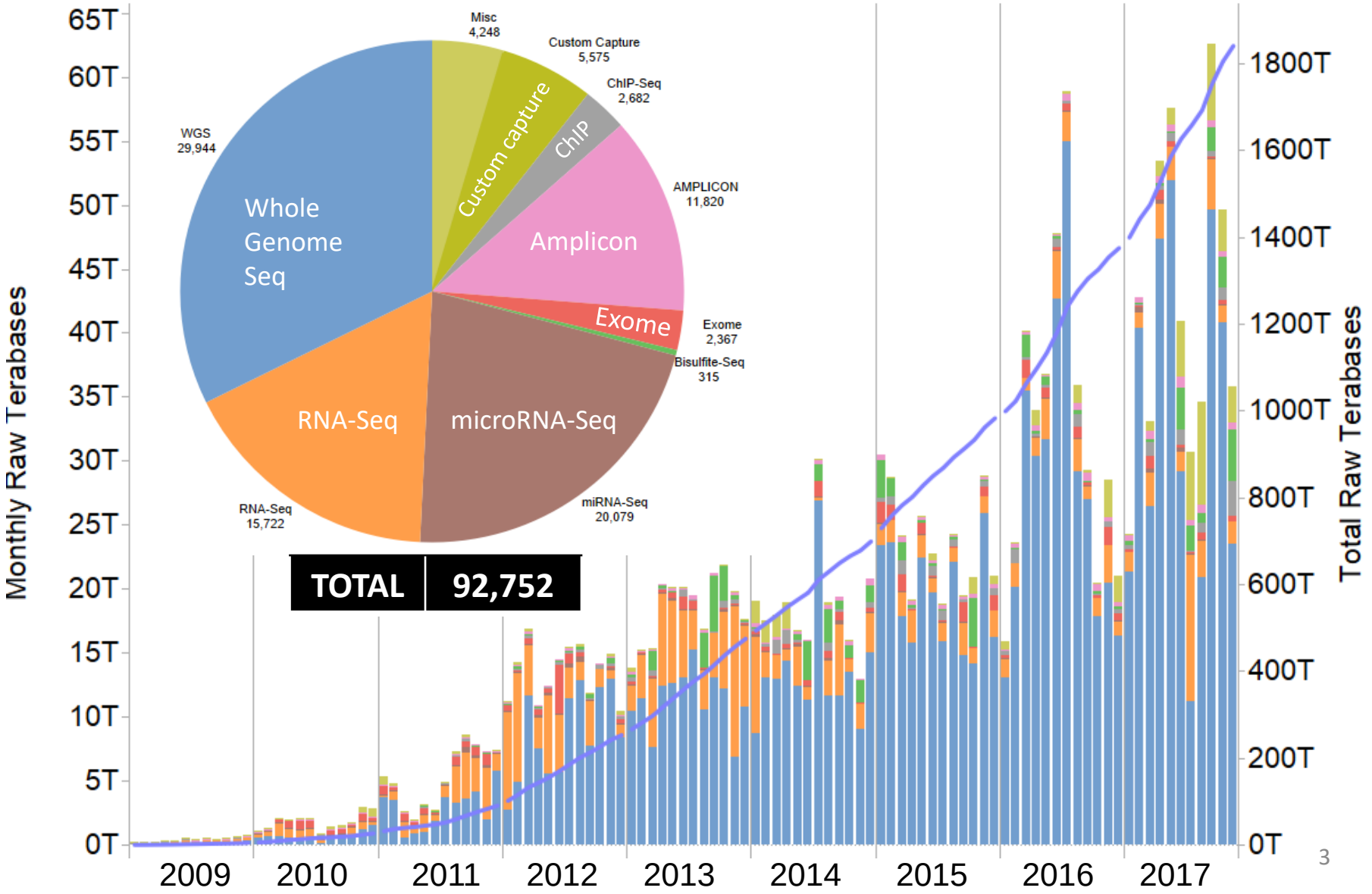
Our mandate is to advance knowledge about cancer and other diseases and to use our technologies to improve health through disease prevention diagnosis and therapeutic approaches.

As a Process Development Coordinator I help ensure our laboratory and analytical approaches are providing the best possible results.

BC Cancer Agency
CARE & RESEARCH
An agency of the Provincial Health Services Authority

Libraries constructed for Illumina sequencing

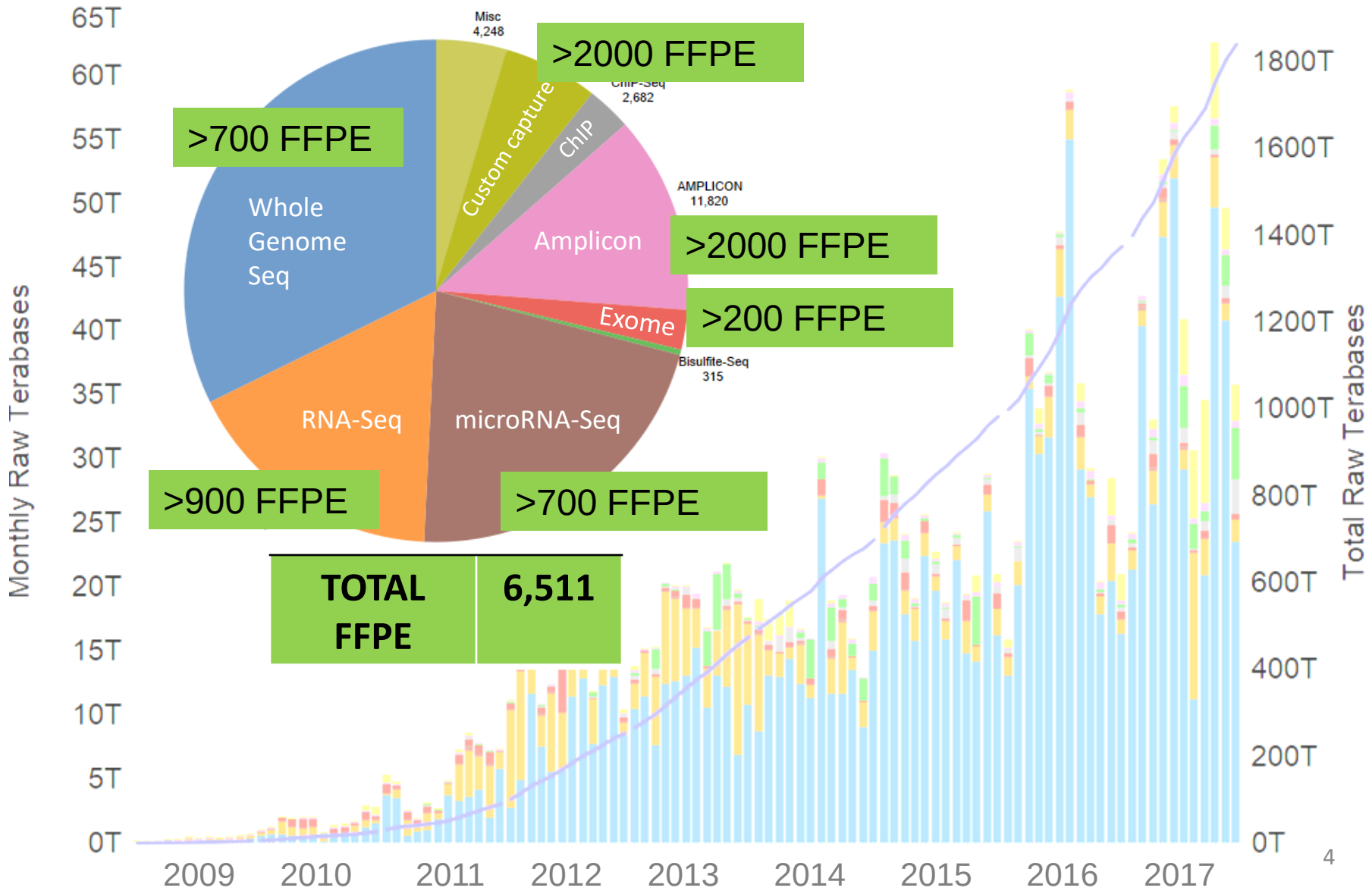
CANADA'S MICHAEL SMITH GENOME SCIENCES CENTRE



BC Cancer Agency
CARE & RESEARCH
An agency of the Provincial Health Services Authority

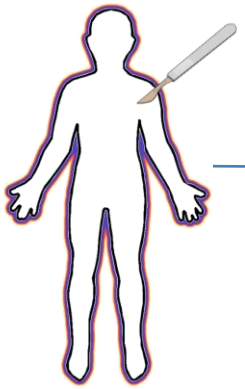
Libraries constructed for Illumina sequencing

CANADA'S MICHAEL SMITH GENOME SCIENCES CENTRE

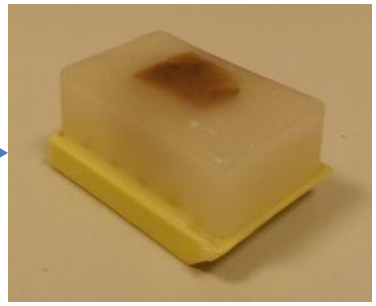


FFPE

- Formalin-Fixed, Paraffin-Embedded



Up to 24 hours



•Can be stored at room temperature.
•Often legally required to be kept for years.

Pros:

- Formalin treatment helps with histological assays
  - Preserves tissue from degradation
  - Holds structure of organelles and cells
- FFPE samples often have well curated clinical information

FFPE Nucleic Acids

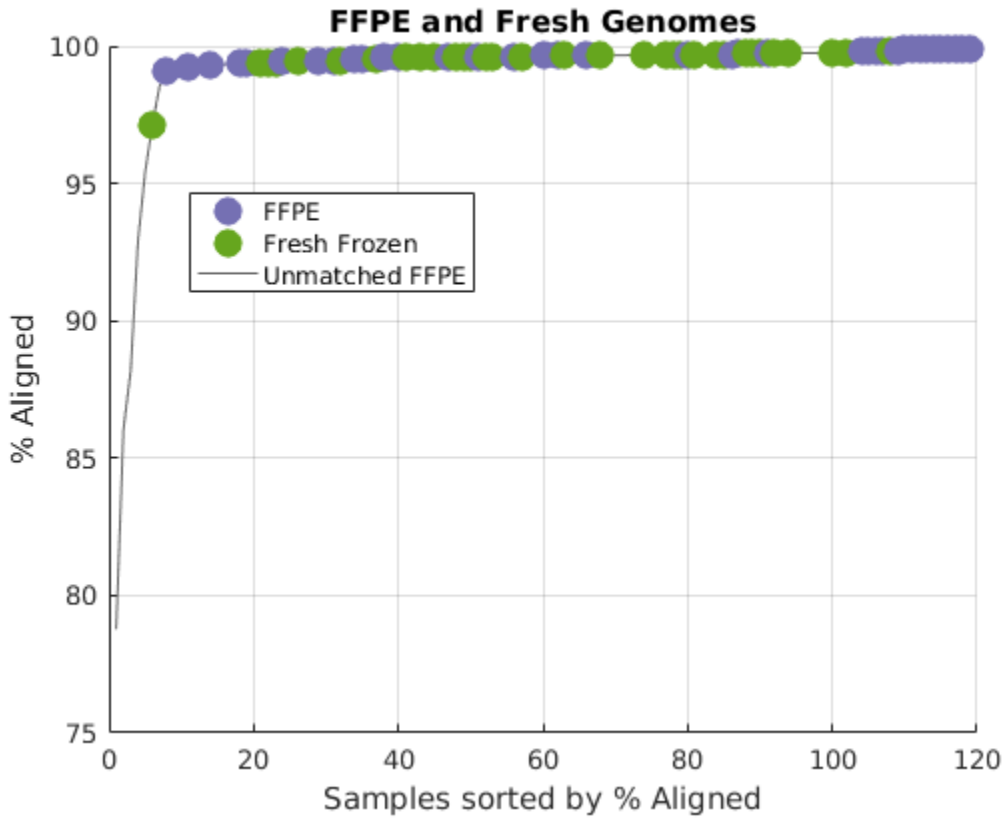
Cons:

- DNA and RNA can be degraded, especially when compared to fresh frozen counterparts.
- Artificial base substitutions can also be present

This gel shows DNA sizes extracted from fresh samples (lanes 3-6,8), and FFPE samples (lanes 1,2,7,9,10)

Lane 2 has high molecular weight, but low yield

FFPE Genome Sequencing

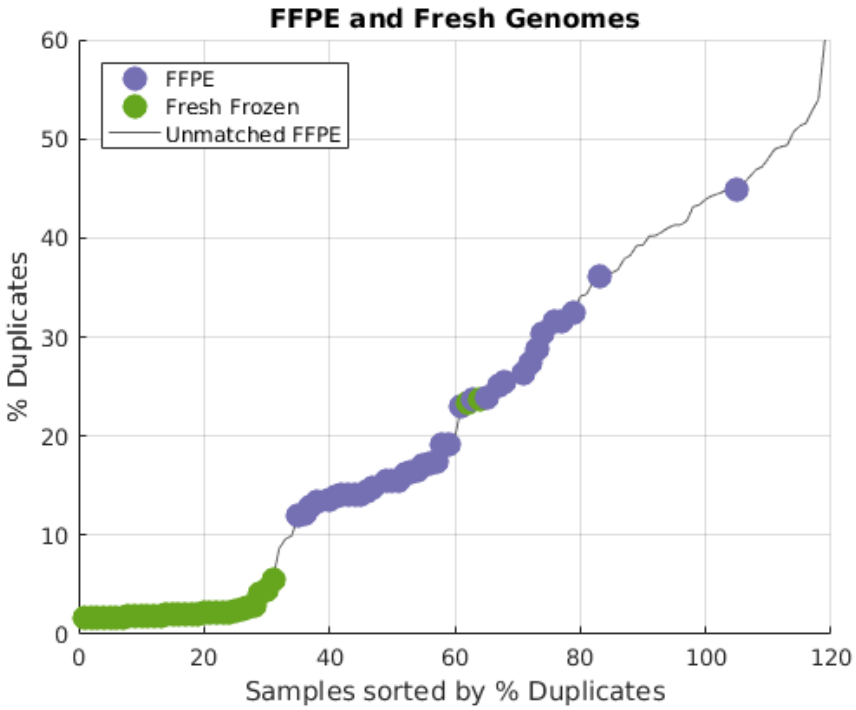


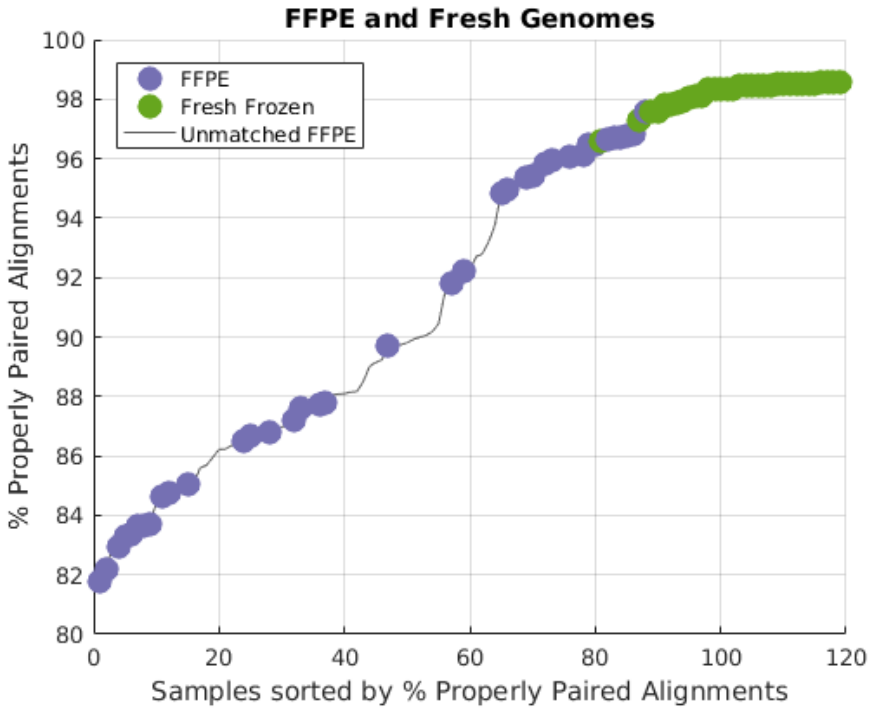
High alignment rates are possible from FFPE and fresh samples.

The samples with lower alignment rates occurred when using sample tracking spike-in plasmid DNA.

Accurate library quantification is helpful.

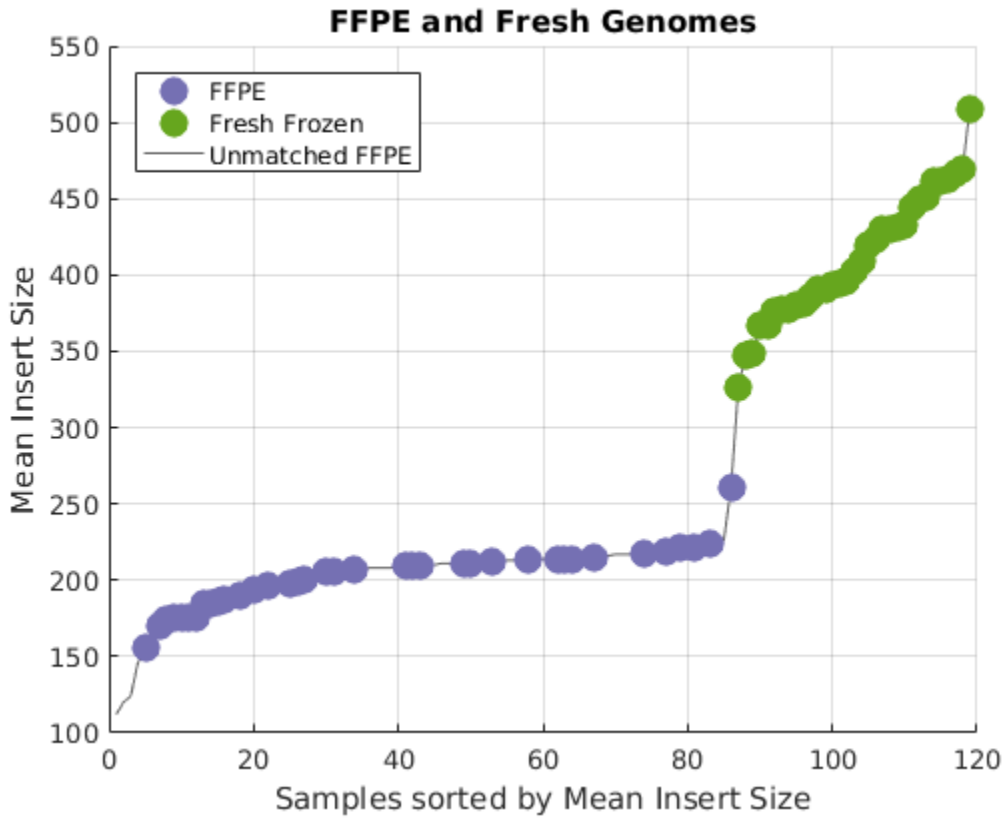
FFPE WGS





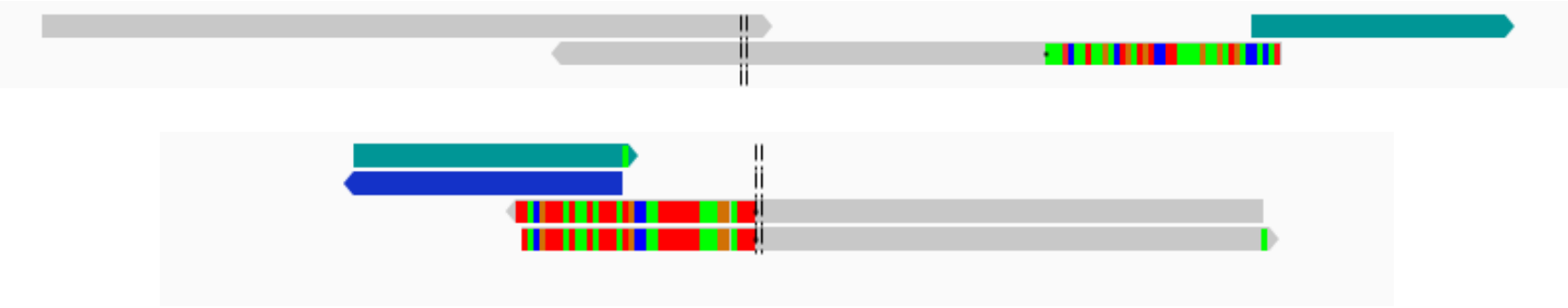
- PCR is usually applied before sequencing. This and frequently limiting DNA amounts often result in higher duplicate rates in FFPE libraries.
- Chimeric fragments (non-properly paired reads) are also more common in FFPE data.

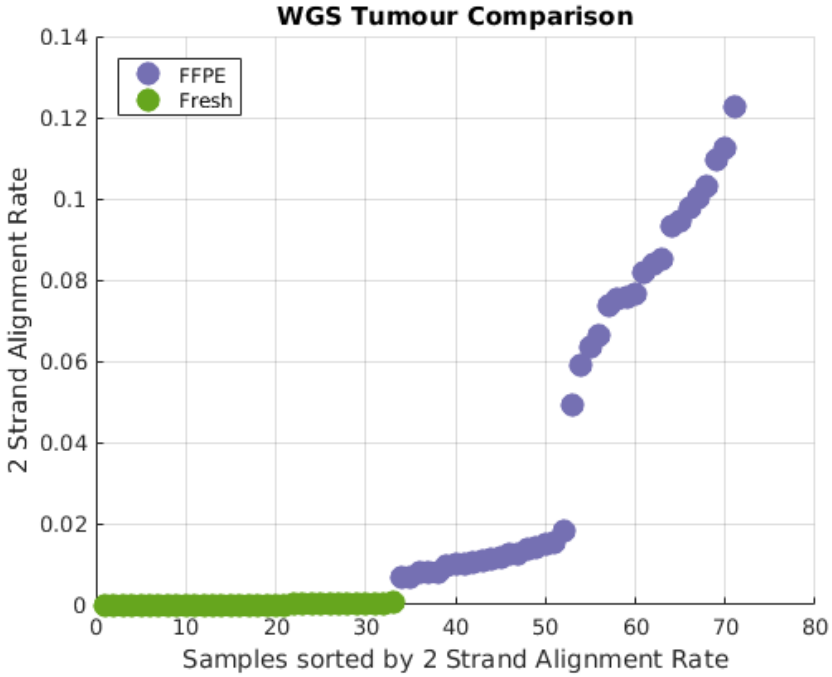
FFPE WGS



Smaller insert sizes in FFPE can:
1.) increase the adapter sequence seen in the reads
2.) create overlapped alignments of reads from a single fragment of DNA

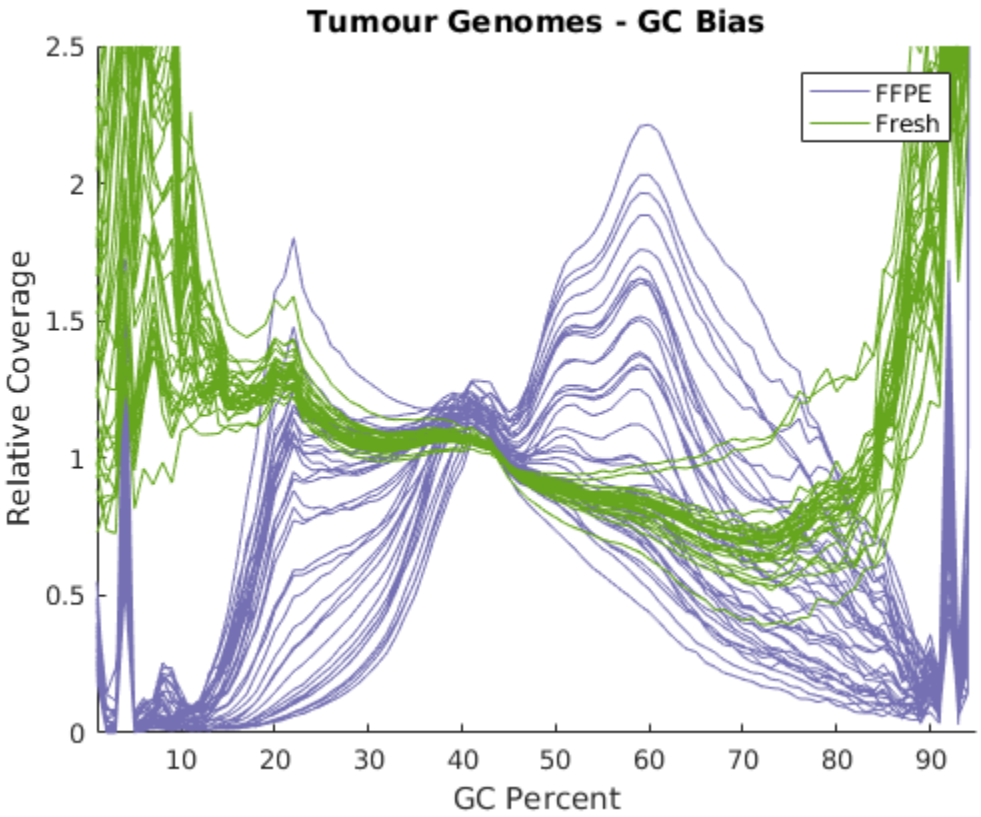
FFPE 2strand





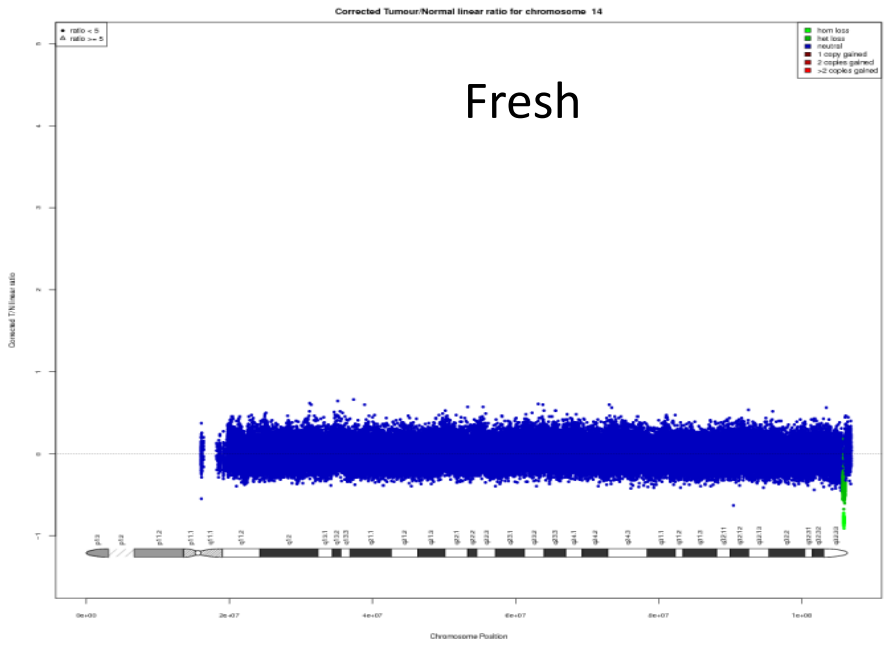
These artifacts are relatively unique to FFPE samples, but can be attenuated by using specific extraction/library construction approaches.

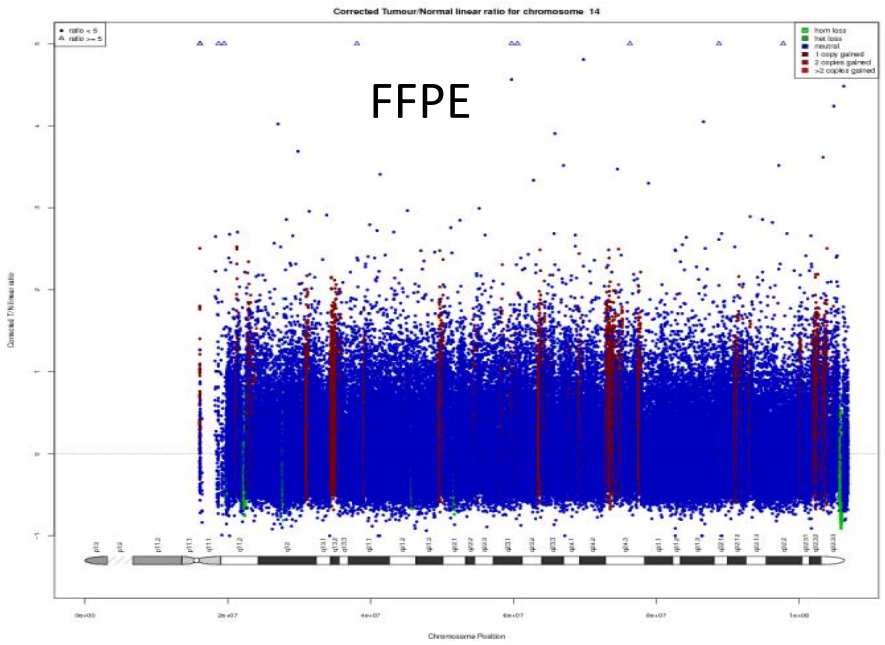
FFPE GC



GC coverage bias is as expected knowing that the FFPE genome library construction includes PCR cycles.

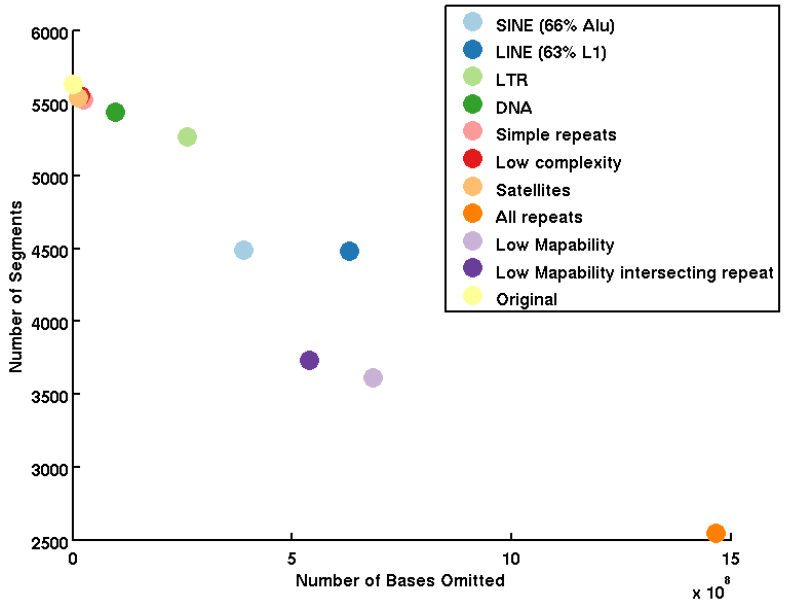
FFPE CNV



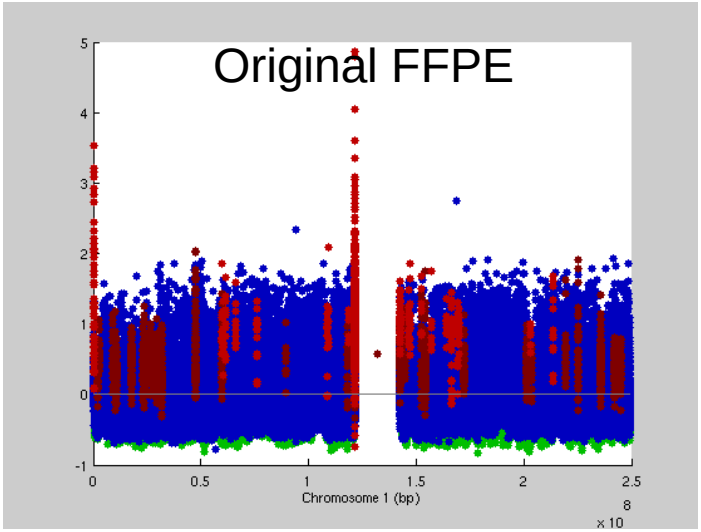


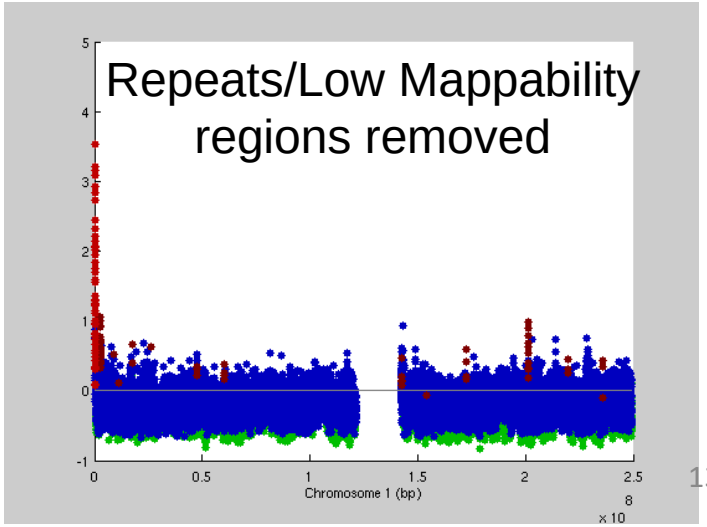
- Many FFPE samples have “noisy” CNV results

FFPE CNV

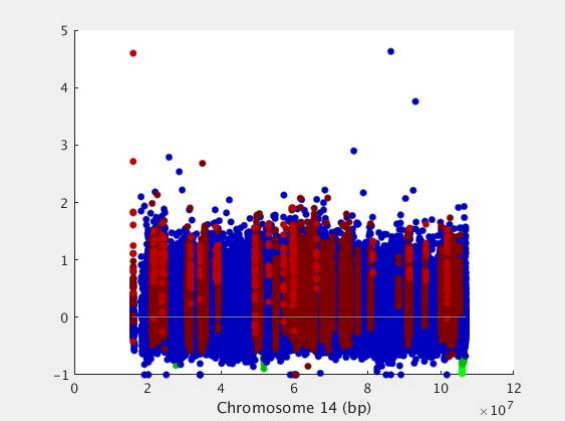


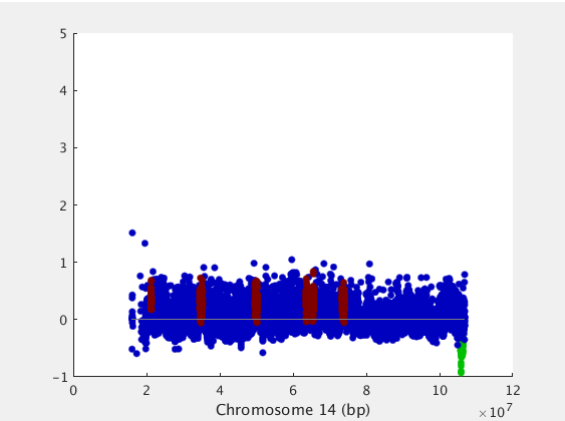
CNV results for some FFPE samples can be significantly cleaned up by omitting parts of the genomes from analysis.

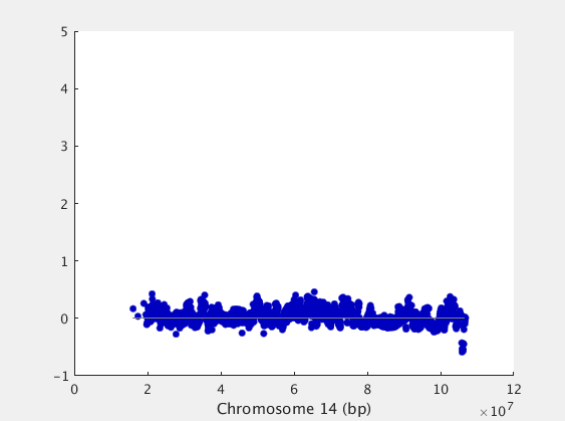


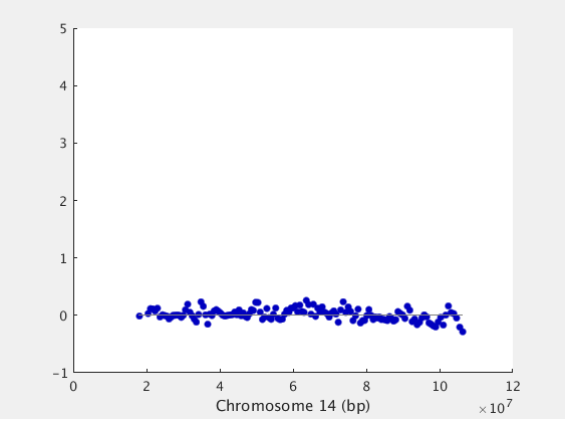


FFPE CNV









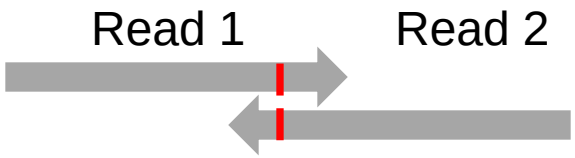
The use of smoothing can also improve the results, but at a cost of resolution.

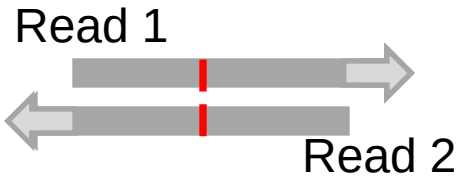
Here, the green (copy number loss) region is correct, while the other events are false positives.

Canvas* is an example of a tool that addressed FFPE noise with signal processing techniques

**Bioinformatics*, Volume 32, Issue 15, 1 August 2016, Pages 2375–2377,https://doi.org/10.1093/bioinformatics/btw163

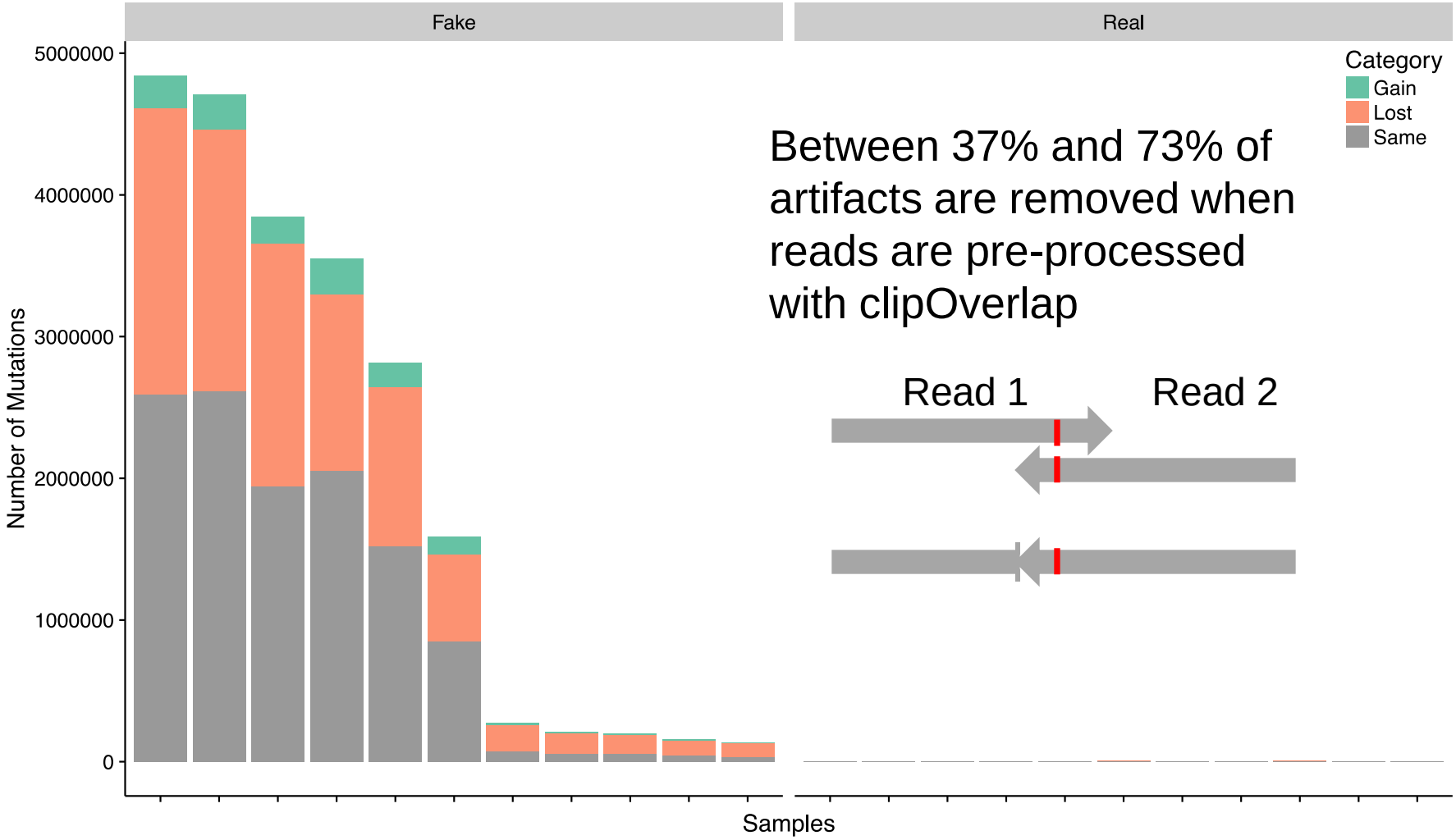
FFPE Variant Call





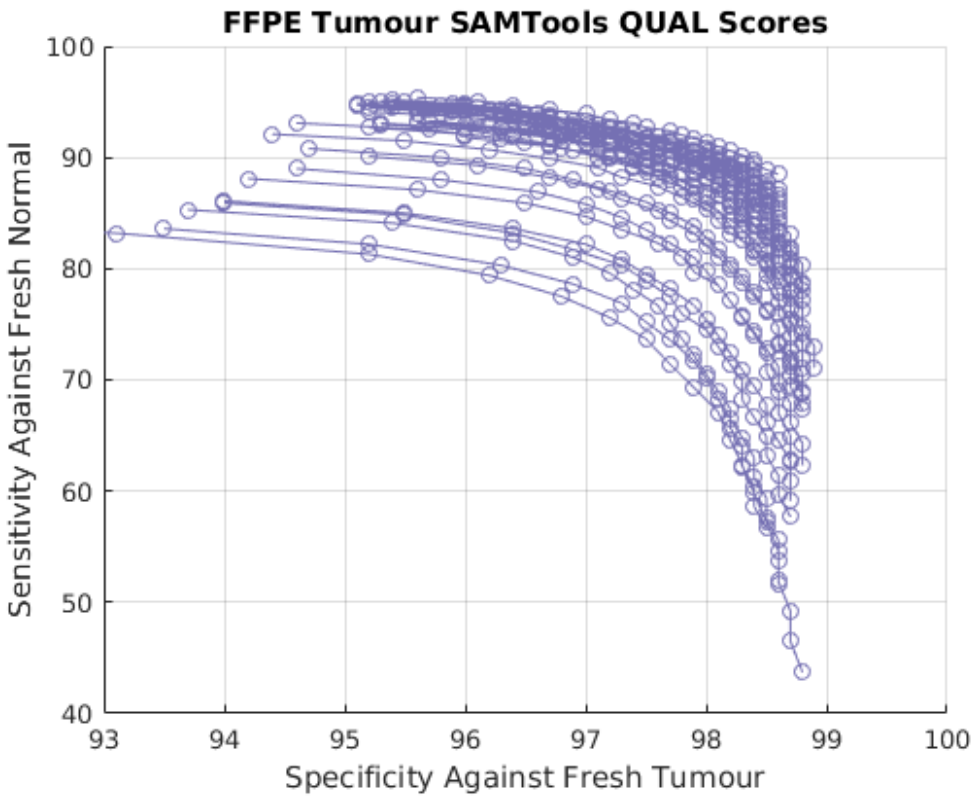
- Frequently, FFPE samples have an increased rate in false positive variant calls. The specific base changes can differ depending on a sample's history or the application of "repair reagents".
- Small insert sizes in FFPE libraries can introduce 2 observations of an error, causing false positive calls.

FFPE Variant Call



Between 37% and 73% of artifacts are removed when reads are pre-processed with clipOverlap

FFPE Variant Call



Single sample FFPE SAMTools variant calls can be tuned by adjusting the required quality score to yield higher sensitivity or specificity.

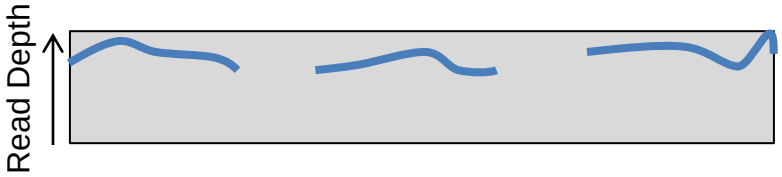
FFPE RNA

- RNA can be sequenced from FFPE samples.
  - Often starting from lower RNA amounts than fresh samples
  - FFPE RNA is often more degraded than RNA from fresh samples

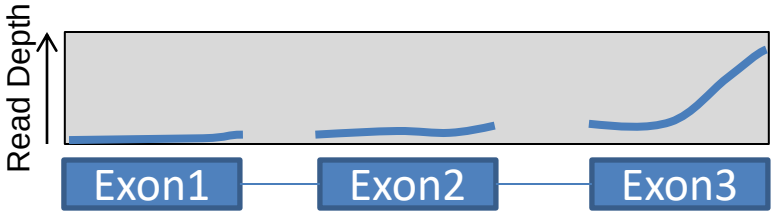
FFPE RNA

TAGGTATACGCCCTGAGTGGATGACCCGCTAGCTAGAAAAAAAAAA
TAGGTATACGCCCTGAGTGGATGACCCGCTAGCTAGAAAAAAAAAA
TAGGTATACGCCCTGAGTGGATGACCCGCTAGCTAGAAAAAAAAAA
TAGGTATACGCCCTGAGTGGATGACCCGCTAGCTAGAAAAAAAAAA

Intact RNA



Intact RNA captured using the polyA tail yields even read coverage for the whole transcript



Degraded RNA captured using the polyA tail yields high coverage of 3' end

TAGGTATACGCCCTGAGTGGATGACCCGCTAGCTAGAAAAAAAAAA
TAGGTATACGCCCTGAGTGGATGACCCGCTAGCTAGAAAAAAAAAA
TAGGTATACGCCCTGAGTGGATGACCCGCTAGCTAGAAAAAAAAAA
TAGGTATACGCCCTGAGTGGATGACCCGCTAGCTAGAAAAAAAAAA

Degraded RNA

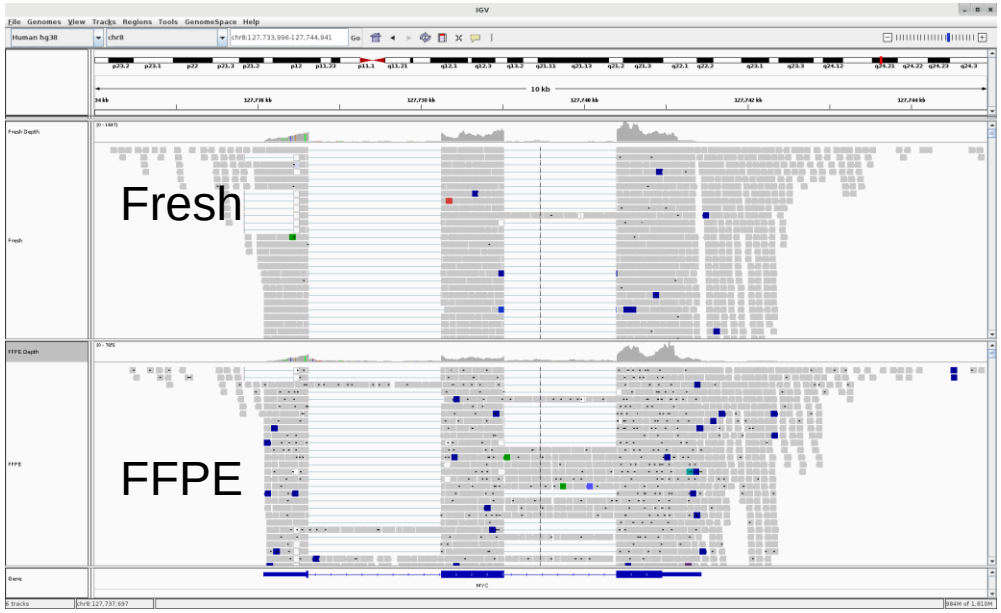
-> Use ribosomal depletion

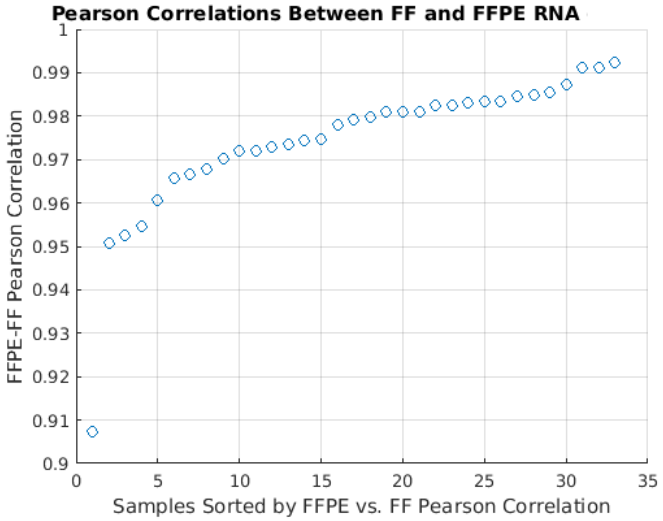
FFPE RNA

– Random primed cDNA synthesis in RBD RNA-seq counters 3' end bias otherwise seen in polyA capture of degraded transcripts.

Fresh Samples	FFPE samples
polyA selection	Ribosomal depletion
•Targets messenger RNA •High Exon/Intron ratios	•Contains coding and non-coding RNAs •Increased read diversity, including: •Intergenic content •Intronic content

FFPE RNA

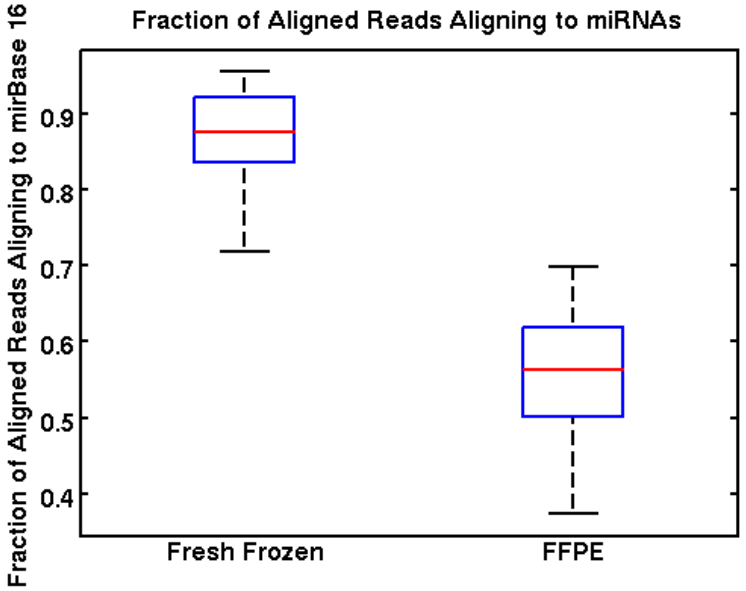


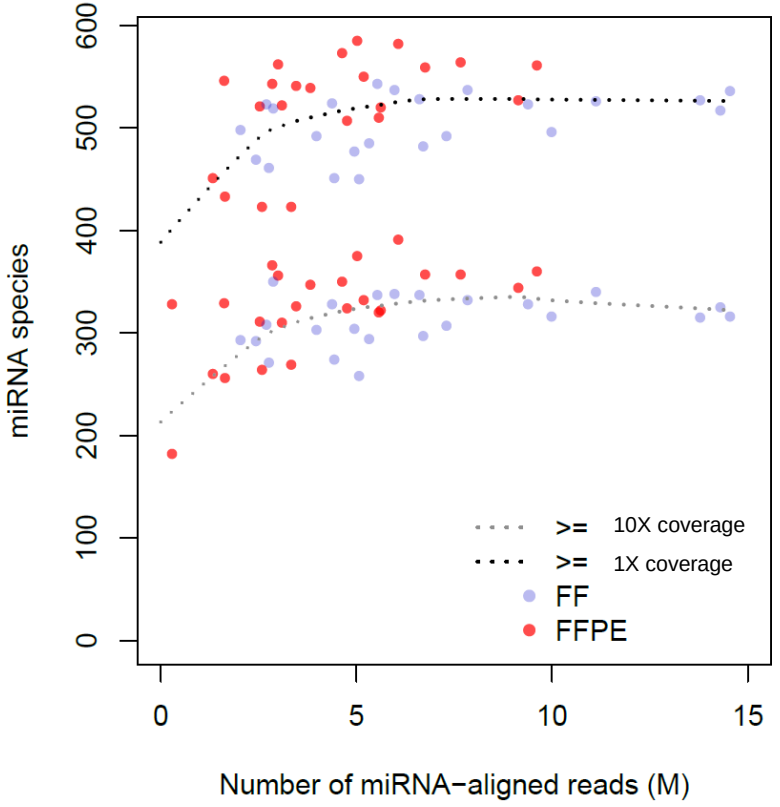


The FFPE RNA library has much higher intronic content creating a lower exon/intron ratio

Without extra analysis Fresh and FFPE samples from the same sources correlate well

FFPE miRNA





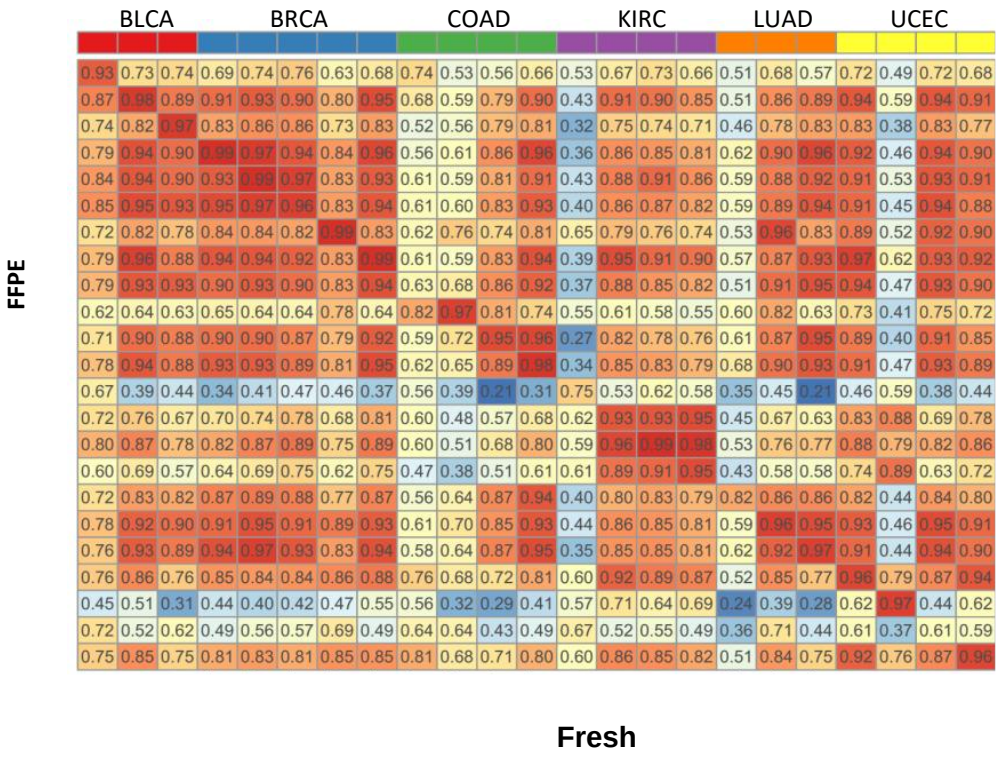
miRNA libraries can be constructed from FFPE sources. These libraries usually have:

- Lower fraction of reads aligning to annotated miRNAs
- Higher detected diversity per number of reads

BC Cancer Agency
CARE & RESEARCH
An agency of the Provincial Health Services Authority

CANADA'S MICHAEL SMITH GENOME SCIENCES CENTRE

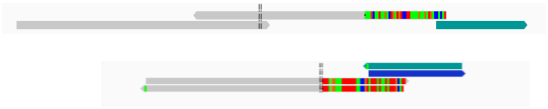
FFPE miRNA cluster



Many FF/FFPE pairs are each other's best correlate
16/23 pairs cluster together

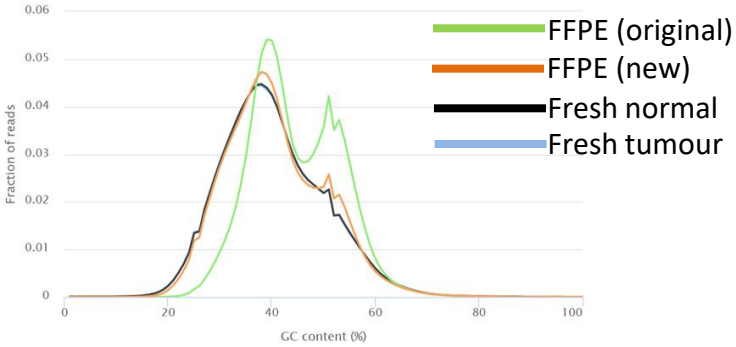
Wet Lab Improvements

2-Strand Artifact:



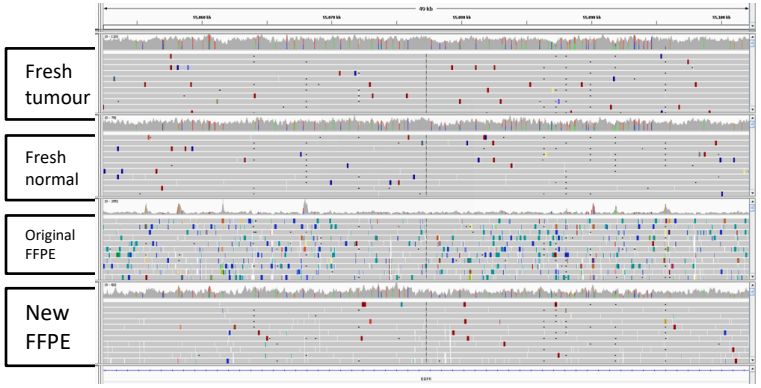
Micro inversions in FFPE data like those shown above have been reduced by >90%

Read GC content:



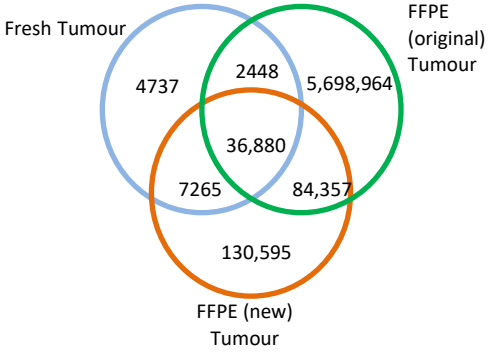
Recent developments have improved the GC content of FFPE libraries to approximate that of libraries from fresh frozen sources.

Uneven Coverage and Poor Pairing Rates:



IGV screenshots of 4 libraries from the same individual (2 fresh, 2 FFPE), show a significant improvement in coverage evenness and proper pairing of reads in the new FFPE data.

Somatic Variant Calls:



False positive variant calls are reduced when using newer approaches

Selected Publications

- Comprehensive miRNA sequence analysis reveals survival differences in diffuse large B-cell lymphoma patients.
  - Lim *et al.* Genome Biol. 2015 Jan 29;16:18. doi: 10.1186/s13059-014-0568-y
- Burkitt Lymphoma Genome Sequencing Project (BLGSP): Integrative Genomic and Transcriptomic Characterization of Burkitt Lymphoma
  - Grande *et al.* ASH Abstract 2017.
- Automated high throughput nucleic acid purification from formalin-fixed paraffin-embedded tissue samples for next generation sequence analysis.
  - Haile *et al.* PLoS One. 2017 Jun 1;12(6):e0178706. doi: 10.1371/journal.pone.0178706. eCollection 2017.
- Comprehensive characterization of genomic, transcriptomic and epigenomic artifacts introduced in formalin-fixed, paraffin-embedded tissues.
  - Zmuda *et al.* Submitted, Oct. 2017

Summary

- FFPE samples may be highly variable within a set
- Nucleic Acids derived from FFPE can be processed in a high throughput fashion to create sequenceable DNA and RNA libraries.
- Data from FFPE samples can be a beneficial contribution to research activities