# New Method Enables Analysis of Preserved Cancer Specimens With Exome Sequencing

A novel QC approach and the Nextera® Rapid Capture Exome Kit enable researchers to sequence degraded DNA from FFPE tumor samples successfully.

## Introduction

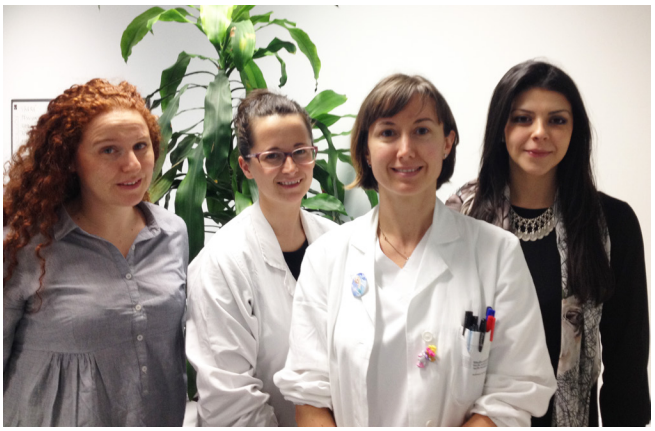
The specific causes of cancer progression have long baffled scientists and physicians. Repositories are filled with historical tumor samples collected from patients over the years that might hold the answers. Unfortunately, while the process used to preserve tumor samples is successful in saving them for decades, it often introduces DNA analysis challenges. Formalin-fixed, paraffin-embedded (FFPE) tumor samples usually contain degraded DNA, which is difficult to amplify for next-generation sequencing (NGS).

Led by Annalisa Astolfi, PhD, researchers at the Centro Interdipartimentale di Ricerche sul Cancro "Giorgio Prodi" (CIRC) at the University of Bologna in Italy are developing new ways to retrieve the information preserved in FFPE tumors. They hope that this information will provide a better understanding of the genetic factors that cause cancer. Dr. Astolfi and her colleagues began using microarrays to analyze gene expression in solid tumors and leukemia, but later incorporated NGS into their analysis methods. "The advent of NGS completely changed our studies," Dr. Astolfi said. "We can now investigate tumor samples more deeply with sequencing, enabling us to search for novel somatic mutations in rare tumors. We don't have much information about the somatic events that drive tumor onset and progression in these cancers."

While the CIRC team started sequencing the exomes of fresh-frozen (FF) tumor samples, their research on gastrointestinal stromal tumors (GIST) led them to explore FFPE samples. "Traditionally, GIST has been characterized by mutually exclusive *KIT* or *PDGFRA* mutations, and recent findings have shown that inactivating *SDH* mutations are also associated with GIST,"[1,2] Dr. Astolfi added. "GIST studies focus on the proportion of individuals with no known genetic marker for the disease, and on the accompanying genetic lesions that characterize disease progression and resistance to therapy. We're investigating other mutations that might contribute to GIST."[3]

To address the challenge of sequencing degraded DNA, Dr. Astolfi and colleagues developed a quality check (QC) method for FFPE samples. This method allows them to assess whether a particular sample will yield high-quality data and whether to proceed with sequencing. Their QC step uses a modification of the Random Amplified Polymorphic DNA (RAPD) assay, a well-known method of DNA fingerprinting.[4] It produces an unbiased representation of DNA quality through random amplification of multiple loci throughout the genome. "If we can select samples of high quality, then we can obtain high-quality data that are comparable to FF samples," Dr. Astolfi said.

This application note describes the QC method, DNA analysis of FF and FFPE GIST, and data comparison with matched normal tissue.



**Valentina Indio, PhD**; **Milena Urbini, MSc**; **Annalisa Astolfi, PhD**; and **Chiara Giusy Genovese, MSc** (from left to right) at CIRC are developing new methods for investigating and understanding cancer progression.

## Sample Collection and DNA Extraction

The CIRC team collected clinical tumor samples, using established methods to preserve a portion of the samples as FF. The remaining portion of tumor specimens were fixed in 10% formalin solution for 6–72 hours, and then dehydrated and embedded in paraffin. FF samples were extracted with the QIAamp DNA Mini Kit (QIAGEN). FFPE slides were manually macrodissected by superimposing an H and E stained slide, then extracted with the QIAamp DNA Micro Kit (QIAGEN). DNA concentration was determined using both spectrophotometric (NanoDrop, Thermo Scientific) and fluorometric (Quant-IT PicoGreen dsDNA Assay Kit, Life Technologies) methods.*

> "The advent of NGS completely changed our studies. We can now investigate FF and FFPE tumor samples more deeply with sequencing, enabling us to search for novel somatic mutations in rare tumors."

* Illumina recommends using fluorometric quantification to determine DNA concentration with Illumina products.

For Research Use Only. Not for use in diagnostic procedures.

## Quality Check on FFPE Samples

The CIRC team developed a QC step, using a modified RAPD method to assess FFPE quality (Figure 1). Briefly, 5 ng of DNA from FF and FFPE DNA were amplified with RAPD primers, KAPA2G Fast HotStart Taq Polymerase (Kapa Biosystems), $MgCl_2$, and dNTPs. PCR was performed for 40 cycles.

To demonstrate concordance, the CIRC team also performed a QC check using an established, gene-specific, quantitative PCR method (Kapa Biosystems). The results of the QC step were in concordance with the QC results following the modified RAPD method.

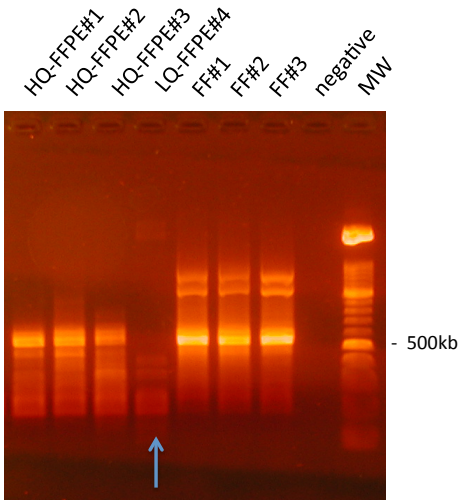



**Figure 1: Amplification Results of QC Step**—Gel electrophoresis of RAPD-amplified DNA demonstrated that high-quality FFPE samples show a fingerprint pattern that resembles the profile of FF samples, with the presence of a prevalent, bright, 500 bp fragment as the highest amplicon size. Conversely, low-quality FFPE samples (denoted by the blue arrow) amplify only short fragments, with faint bands of less than 200 bp.

## Library Preparation and Exome Sequencing

The team used the Nextera Rapid Capture Exome (NRCE) Kit to prepare DNA libraries from FF tumor tissue, FFPE tumor tissue, and matched normal peripheral blood. FFPE samples were chosen to represent 3 high-quality samples (HQ-FFPE) and 1 low-quality sample (LQ-FFPE). Starting with 100 ng of genomic DNA, they followed the Nextera Rapid Capture protocol to tagment (tag and fragment) DNA, denature libraries to single-stranded DNA, hybridize to biotinylated probes, and pool libraries from FF and FFPE samples.

"Our experience with the NRCE Kit has been very positive," said Dr. Astolfi. "We wanted to develop a protocol that was easy to implement and that could be performed in any biomolecular laboratory. NRCE possesses its own DNA tagmentation reaction and does not require initial DNA fragmentation by Covaris or another method. The ability to pool a different number of samples per reaction also makes NRCE flexible for users with distinct needs. It's important to note that depending on the average size of the fragments, labs will need to determine the amount of input DNA and the concentration of the transposon enzyme to use."

### Modified RAPD Method for FFPE QC

#### PCR Mix (25 µl total volume)

| | |
|---|---|
| 2.5 µl | 10X buffer |
| 2 µl | 25 mM MgCl2 |
| 0.5 µl | 10 mM dNTPs |
| 1.25 µl | 10 µM forward primer (5'-AATCGGGCTG-3') |
| 1.25 µl | 10 µM reverse primer (5'-GAAACGGGTG-3') |
| 0.1µl | Taq polymerase |

#### Thermal Cycler Protocol

| | | | |
|---|---|---|---|
| 95 ºC | | 2 minutes | |
| | 95 ºC | 15 seconds | 40 cycles |
| | 37 ºC | 15 seconds | |
| | 72 ºC | 1 minute | |
| 72 ºC | | 30 seconds | |

Library preparation yielded sufficient material for both FF and HQ-FFPE samples, even though average library size was smaller (Figure 2). For LQ-FFPE samples, the team pooled 2 different reactions to yield 500 ng of tagmented DNA.

> "Our experience with the NRCE Kit has been very positive."

Following library preparation, the team sequenced the libraries in a 2 × 100 bp whole-exome sequencing run on the HiScanSQ™ System. Sequencing generated an average of 57 million reads for FF and 56 million reads for HQ-FFPE samples. All other sequencing parameters were comparable between FF and HQ-FFPE samples (Table 1). Conversely, LQ-FFPE failed to meet sequencing thresholds and quality parameters, with a much lower data yield (14 million reads) and low average coverage.

## Analysis and Results

### Alignment and Variant Calling

After quality trimming and adapter removal, sequence reads were mapped to the human reference genome hg19 with BWA software (Figure 3). Alignments were processed using SAMtools to remove PCR duplicates and using GATK to perform local realignment around the indel position, base quality score recalibration, and indel calling. Variant calling was performed with MuTect to identify single nucleotide variants (SNVs). Only variants that were covered at > 10× depth and mapped onto the 37 Mb NRCE target region were analyzed further. "We observed that the quality of the variant call is lower in LQ-FFPE samples," said Valentina Indio, PhD, bioinformatics scientist at CIRC. "In contrast, the number of variant calls and the quality of base calls in HQ-FFPE samples are similar to FF samples."

For Research Use Only. Not for use in diagnostic procedures.

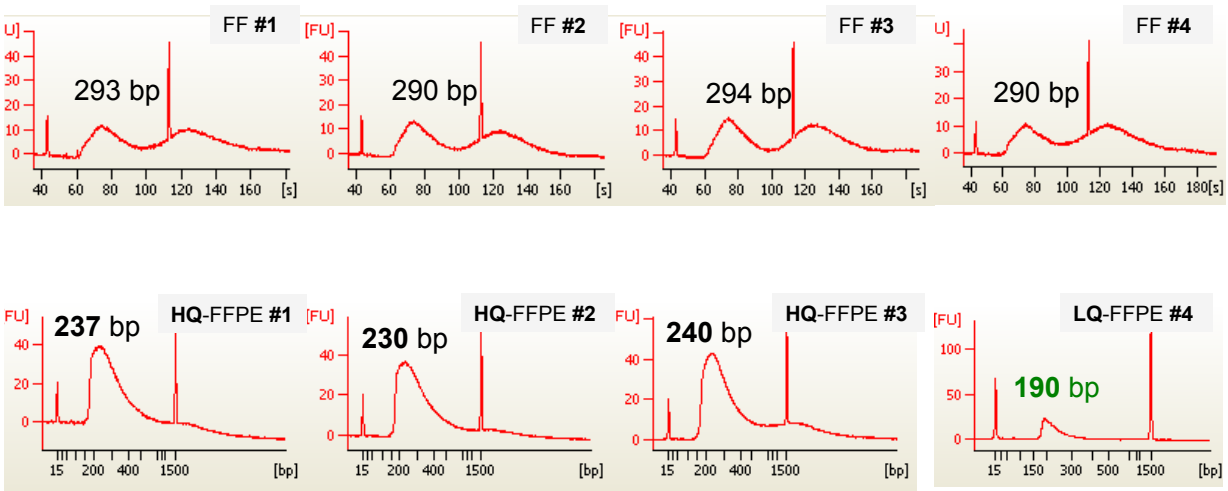



**Figure 2: Library Preparation Results**—Library preparation yielded sufficient material for both FF and HQ-FFPE samples, even though average library size was slightly smaller for HQ-FFPE samples (235 bp vs 292 bp). In contrast, the LQ-FFPE sample yielded a significantly smaller library. Library profiles were generated using the Bioanalyzer (Agilent Technologies).

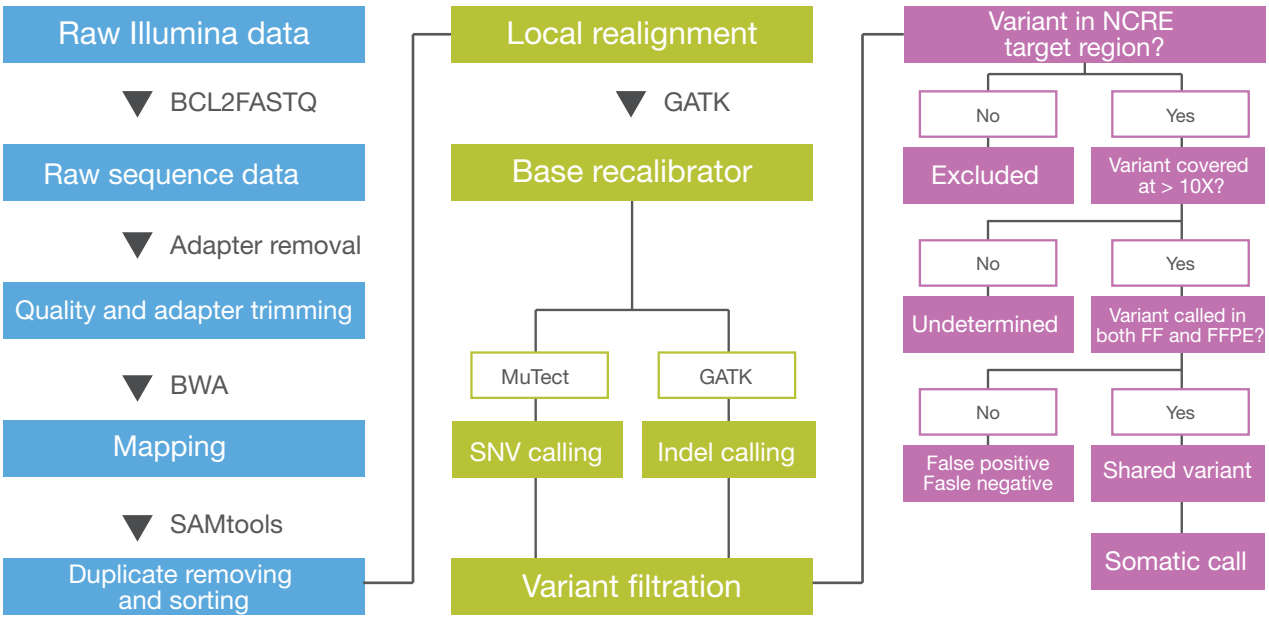



**Figure 3: Analysis Workflow**—The CIRC researchers used an analysis workflow that includes alignment and variant calling, comparison of FF and FFPE results to identify false positives and false negatives, and somatic variant identification.

## Comparison of Sequencing Results for FF and FFPE Samples

The CIRC team analyzed the resulting data set to estimate the agreement between FFPE and FF data. Variants that appeared in both FF and FFPE samples were designated "shared variants."

"HQ-FFPE samples showed a shared variant concordance of 94–96%, while LQ-FFPE samples have a shared variant concordance of about 50%," Dr. Indio noted.

## Somatic Variant Identification

"To highlight somatic variants, we filtered the concordant variants against databases of human variability, such as dbSNP[5] and 1000 Genomes,"[6] explained Dr. Indio. "We excluded the variants that are defined as polymorphisms (1% incidence in the normal population), defining the remainder as somatic variants if they were absent in the normal counterpart. We then selected the nonsynonymous mutations and indels that lead to frameshift mutations. To prioritize the variants correlated to disease, we used software to predict the effects of the mutation on protein structure."

For Research Use Only. Not for use in diagnostic procedures.

Table 1: Sequencing Results

| Sample | Bases > Q30 | Trimmed Bases | PCR Duplicates | Mapped Reads | Off-Target Reads | Average Coverage |
|---|---|---|---|---|---|---|
| FF #1 | 83.6% | 8.1% | 7.8% | 91.1% | 51.5% | 71× |
| FF #2 | 84.0% | 7.7% | 7.6% | 91.4% | 52.7% | 59× |
| FF #3 | 84.4% | 9.3% | 7.5% | 91.2% | 52.0% | 61× |
| FF #4 | 84.0% | 8.3% | 7.1% | 91.9% | 50.8% | 55× |
| HQ-FFPE #1 | 84.7% | 9.4% | 10.3% | 88.4% | 45.7% | 65× |
| HQ-FFPE #2 | 85.0% | 9.3% | 10.4% | 88.4% | 47.1% | 58× |
| HQ-FFPE #3 | 85.9% | 9.1% | 10.0% | 89.0% | 45.5% | 77× |
| LQ-FFPE #4 | 82.0% | 20.6% | 12.0% | 86.7% | 38.8% | 17× |

"From this analysis, we identified 4, 13, 25, and 27 somatic mutations in the 4 FF GIST samples. We defined the somatic mutations as shared variants if they were also present in FFPE samples (3, 13, 24, and 14, respectively)," noted Dr. Indio. "False-positive variants (present only in the FFPE samples) were negligible in HQ-FFPE. In contrast, in LQ-FFPE, false positives represented > 40% of the identified variants. Most of the false-positive calls in the LQ-FFPE were C to T and G to A substitutions, an expected result of the cytosine deamination due to formalin fixation. In the HQ-FFPE samples, the distribution of substitution type is uniform and comparable to FF samples."

> "We were surprised to perform exome sequencing on FFPE samples so successfully, with comparable data for FF and FFPE samples."

## Conclusions

By assessing FFPE sample quality and only choosing high-quality samples for sequencing, Dr. Astolfi's team achieved exome sequencing results with FFPE samples that were comparable to FF samples. "We recommend that researchers not proceed with exome enrichment and sequencing of LQ-FFPE samples, and instead select high-quality samples to generate more useful data," Dr. Astolfi said. "For example, in our GIST database, we estimated that around 40% of FFPE specimens are of high quality. We were surprised to perform exome sequencing on FFPE samples so successfully, with the data generated comparable to that achieved with FF samples. By using this method, we were able to include more GIST samples for which the underlying tumorigenic event was not known. We hope to identify at least 1 gene that seems to be related to the pathogenesis of these tumors."

The CIRC researchers plan to apply their FFPE exome sequencing method to more tumor samples from different cancer types. "We want to expand this analysis to more cohorts of patients with detailed clinical data," Dr. Astolfi said. "We're trying to understand how the genetic basis of tumors might influence reactions to therapeutic drugs, and how the molecular background of rare tumors differs from more common tumors."

## References


1. Pantaleo MA, Astolfi A, Indio V, et al. *SDHA* loss-of-function mutations in *KIT-PDGFRA* wild-type gastrointestinal stromal tumors identified by massively parallel sequencing. *J Natl Cancer Inst.* 2011;103:983-987.
2. Pantaleo MA, Astolfi A, Urbini M, et al. Analysis of all subunits, SDHA, SDHB, SDHC, SDHD, of the succinate dehydrogenase complex in KIT/PDGFRA wild-type GIST. *Eur J Hum Genet.* 2014;22:32-39.
3. Nannini M, Astolfi A, Urbini M, et al. Integrated genomic study of quadruple-WT GIST (*KIT/PDGFRA/SDH/RAS* pathway wild-type GIST). *BMC Cancer.* 2014;14:685.
4. Siwoski A, Ishkanian A, Garnis C, Zhang L, Rosin M, Lam WL. An efficient method for the assessment of DNA quality of archival microdissected specimens. *Mod Pathol.* 2002;15:889-892.
5. dbSNP (www.ncbi.nlm.nih.gov/SNP)
6. 1000 Genomes (www.1000genomes.org)


**Illumina** • 1.800.809.4566 toll-free (US) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com

**For Research Use Only. Not for use in diagnostic procedures.**

© 2015 Illumina, Inc. All rights reserved. Illumina, HiScanSQ, Nextera, and the pumpkin orange color are trademarks of Illumina, Inc. and/or its affiliate(s) in the U.S. and/or other countries. All other names, logos, and other trademarks are the property of their respective owners.
Pub. No. 370-2015-001 Current as of 16 April 2015

illumina®

AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAAC
TCAACGTACCGTAACGAACGTATCATTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTT
CGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACC
AACGTACCATTAAGAGCTACCGTGCAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAA
AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAAC
GATTACTTGATCCACTGATTCAACGTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAGCAACGACG
CGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGAAAAGAATGATAACAGTA