# illumina

# Analyzing FFPE Samples with Infinium<sup>®</sup> Microarrays

A streamlined workflow for analyzing FFPE samples in GenomeStudio® Data Analysis Software.

### Introduction

Formalin-fixed, paraffin-embedded (FFPE) tissue samples provide an invaluable resource and repository of information that can provide insight into the underlying genetic mechanisms associated with disease. These samples can be useful in biomarker discovery, drug development, and cancer research, yielding high-quality data in whole-genome genotyping studies.

However, the fixation process and storage of FFPE samples frequently leads to nucleic acid degradation and base modification. Because it is inefficiently amplified in the whole-genome amplification step, degraded DNA from FFPE samples generally results in poor performance in array-based genotyping studies. Such samples have traditionally been unusable with the Illumina Infinium assays. To overcome this problem, the Infinium FFPE DNA Restore Kit provides a simple workflow to repair damaged DNA, achieving high-quality genotyping and copy number variation data in Infinium assays. Details of this workflow are described in the data sheet, Infinium FFPE DNA Restoration Solution, Pub. No. 370-2010-009.

Furthermore, analyzing the data generated from FFPE samples using standard cluster-fitting algorithms can be complicated. Often, the performance differences between normal and FFPE samples cause changes in the cluster positions of some markers, leading to inaccurate genotyping and copy number calling. To increase the accuracy of the genotyping software and to analyze data generated from FFPE samples more effectively, special analysis procedures are necessary to account for clustering differences. The most direct method for managing the cluster-fitting algorithm is to identify the differentially performing markers (SNPs) and exclude them from further analysis\*.

This technical note describes a workflow (Figure 1) for repairing damaged DNA and excluding differentially performing FFPE markers from analysis with Illumina GenomeStudio Data Analysis Software. Covering the entire range of Illumina human genotyping microarrays<sup>†</sup> (including iSelect<sup>®</sup> BeadChips) and assays, this method generates high-quality clustering, genotyping, and CNV calling of Infinium Assay data using algorithms in the Genotyping (GT) Module v2011.1/1.9.4<sup>1</sup>.



Figure 1: FFPE Sample QC, Restoration, and

## Simple FFPE Sample QC and Restoration

After DNA extraction using one of many commercially available kits, the Illumina FFPE QC Kit is used to evaluate the quality of prospective FFPE samples, determining if they are useable. The kit provides primers for a real-time PCR assay that can be carried out on most standard real-time PCR instruments.

Samples that pass the quality control evaluation are advanced through the DNA restoration process. Using a proprietary enzymatic reaction, the Infinium FFPE DNA Restore Kit repairs degraded FFPE DNA in preparation for the whole-genome amplification step in Infinium Assays. Generally, over 80% of the FFPE samples that pass the QC assay metric are successfully restored (Figure 2). For optimal assay performance, minor modifications to some of the Infinium workflows are required to increase the effectiveness of the whole-genome amplification step. These modifications are summarized for each assay in the following section.

<sup>\*</sup> Experienced users may find it useful to examine differentially performing markers in more detail and to recluster the markers in question. Information on cluster fitting in GenomeStudio can be found in the technical note, Infinium Genotyping Data Analysis<sup>2</sup>.

<sup>&</sup>lt;sup>†</sup> Because the HumanCytoSNP-FFPE BeadChip has been enabled previously for use with FFPE samples, the method described in this technical note does not apply. In this case, the optimized protocol that accompanies the BeadChip is recommended.



DNA from 130 FFPE samples of varying quality and control DNA (2 ng each) were analyzed in triplicate by real-time PCR. The average Cq value for the control DNA was subtracted from Ct values for FFPE DNA. The resulting Delta Ct values were plotted against the Infinium HD Assay call rates for the samples after restoration. Greater than 80% of samples with a QC value below 5 were successfully genotyped (orange box). Depending on sample origin and age (1–35 years), 35–95% of samples passed QC.

#### Infinium Assay Modifications

For both manual and automated workflows, the amplification step should be performed as instructed in the manual protocol with the following modifications.

- Infinium HD Super Protocol: To prepare the MSA1 plate ("Make MSA1 Plate" step), 8 µl of the repaired sample DNA, instead of the standard 4 µl, are added to the corresponding wells in the MSA1 plate.
- Infinium HTS and HD Ultra Protocols: To prepare the MSA3 plate ("Make MSA3 Plate" step), 8 µl of the repaired sample DNA, instead of the standard 4 µl, are added to the corresponding wells in the MSA3 plate.
- Infinium LCG Protocol: To prepare the MSA6 plate ("Make MSA6 Plate" step), 8 µl of the repaired sample DNA, instead of the standard 4 µl, are added to the corresponding wells in the MSA6 plate.

For all other Infinium assays, whole-genome amplification is performed as instructed in the appropriate Infinium Assay Guides. Following amplification, the remaining protocol steps are performed as recommended using either the manual or automated workflows accordingly.

# **Excluding Differential Markers**

Following the completion of the Infinium Assay, the generated raw data are uploaded to GenomeStudio and evaluated as described in the following sections.

#### **Cluster Normal Samples**

Cluster information for normal samples can be obtained directly from Illumina-supplied cluster files as described in Option #1. When an Illumina-supplied cluster file is unavailable (as with custom arrays), genotyping data from normal samples can be clustered as described in Option #2.

- Option #1: Using the cluster file associated with the array used for genotyping, create a project in GenomeStudio. Select the Index, Name, AA\_R\_Mean, AB\_R\_Mean, and BB\_R\_Mean columns for all markers and export the data to a Microsoft Excel file named with the suffix, "\_Normal-intensities".
- **Option #2:** To cluster normal samples, all FFPE samples are excluded and normal samples are processed using the GenomeStudio automated clustering algorithm. Preferably, at least 100 samples should be analyzed. If 100 samples are not available, fewer samples can be used for clustering, yielding diminishing data quality as the number of samples decreases. After clustering, update the SNP statistics and output data from the SNP table. If applicable, exclude samples that do not pass QC (outliers) using call rate and p10 GC metrics, and recluster the remaining normal samples. After clustering is complete, select the Index, Name, AA\_R\_Mean, AB\_R\_Mean, and BB\_R\_Mean columns for all markers and export the data to an Excel file named with the suffix, "\_Normal-intensities".

#### Table 1: Excel Functions that Identify Differentially -Performing Loci

Differential Metric	Function
AA_R_Mean	abs[AA_R_Mean(Normal)-AA_R_Mean(FFPE)]
AB_R_Mean	abs[AB_R_Mean(Normal)-AB_R_Mean(FFPE)]
BB_R_Mean	abs[BB_R_Mean(Normal)-BB_R_Mean(FFPE)]

Results in the Function column of the Excel file correspond to the absolute value for the intensity differential between the AA\_R\_Mean, AB\_R\_Mean, and BB\_R\_Mean values of the normal and FFPE populations.

#### **Cluster FFPE Samples**

To cluster the FFPE samples, all normal samples are excluded and FFPE samples are processed using the GenomeStudio automated clustering algorithm. Preferably, at least 100 samples should be analyzed. If 100 samples are not available, fewer samples can be used for clustering, yielding diminishing data quality as the number of samples decreases. After clustering, update the SNP statistics and output data from the SNP table. If applicable, exclude samples that do not pass QC (outliers) using call rate and p10 GC metrics, and recluster the remaining FFPE samples. After clustering is complete, select the Index, Name, AA\_R\_Mean, AB\_R\_Mean, and BB\_R\_Mean columns for all markers and export the data to an Excel file named with the suffix, "\_FFPE-intensities".

#### **Compare Data Sets**

After the normal and FFPE samples have been clustered in GenomeStudio and the data exported to Excel files, merge the data sets into a single Excel file. Next, compare the absolute values of the differences between samples at each marker. Calculate the AA\_R\_Mean, AB\_R\_Mean, and BB\_R\_Mean differential metrics for each marker and cluster type by the functions defined in Table 1.

#### **Exclude Differentially Performing Markers**

Rows in which any of the three differential metrics are greater than 0.25 are scored as 1. All other rows are scored as 0. The differential 0,1 column can now be imported into the GenomeStudio Genotyping Module where markers defined as 1 are zeroed and excluded from further analysis. Generally, approximately 10–30% of all markers are removed in this manner.

After excluding differentially performing markers, standard cluster fitting in GenomeStudio can proceed to generate high-quality clustering, genotyping, and CNV data as described in the technical note, Infinium Genotyping Data Analysis<sup>2</sup>.

#### Summary

Archival FFPE samples hold an abundance of invaluable information for many genomic studies, yet they often perform poorly in arraybased assays due to highly degraded DNA. The method for repairing damaged DNA and excluding differentially performing markers described in this technical note increases the accuracy of the GenomeStudio software and enables effective data analysis from these samples. When processed in this manner, high-quality clustering, genotyping, and CNV data can be generated from FFPE samples using any of the Infinium Assays.

#### Learn More

For more information about GenomeStudio Software and supporting documentation, visit www.illumina.com/informatics/sequencing-micro-array-data-analysis/genomestudio.ilmn.

To learn more about the Infinium Genotyping Assays, visit www.illumina.com/technology/infinium\_hd\_assay.ilmn.

#### References

- 1. GenomeStudio Software package, Genotyping (GT) Module v2011.1/1.9.4 (www.illumina.com/software/genomestudio\_software.ilmn).
- Infinium Genotyping Data Analysis (www.illumina.com/documents/products/ technotes/technote\_infinium\_genotyping\_data\_analysis.pdf).

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

#### FOR RESEARCH USE ONLY.

© 2014 Illumina, Inc. All rights reserved. Illumina, GenomeStudio, Infinium, iSelect, the pumpkin orange color, and the Genetic Energy streaming bases design are trademarks of Illumina, Inc. in the U.S. and/or other countries. All other names, logos, and other trademarks are the property of their respective owners. Pub. No. 970-2014-001 Current as of 23 April 2014

