The trouble with FFPE: Handling variable performance of FFPE samples through deep coverage exome sequencing and detecting artifactual mutations from FFPE damage in exome data

Maura Costello, Danielle Dionne, Cole Walsh, Yossi Farjoun, Cassie Crawford, Anna Koutoulas, Kara Slowik, Ruchi Munshi, Louisa Walker, Michael Wilson, Teni Dowdell, Chris Friedrich, Danielle Perrin, Sheila Dodge, and Stacey Gabriel

Broad Institute of MIT and Harvard, Genomics Platform, 320 Charles St, Cambridge, MA, 02141, USA

Introduction

Formalin fixed paraffin embedded (FFPE) samples present both financial and analytical challenges for Cancer exome analysis. Here we present data on the unique behavior of FFPE samples based on the >1000 cancer exomes sequenced in 2015 & development data on improvements to FFPE workflows:

• Why it’s especially challenging to bring FFPE samples to high coverage
• Predicting FFPE performance using upfront QCs
• Comparison of the rate of false artifactual mutations in FFPE vs. Non-FFPE

FFPE samples routinely underperform in the exome process

Data for 294 FFPE exomes and 368 exomes derived from other sources processed in 2015 through our modified shearing-based Illumina Rapid Capture Exome process are compared below.

FFPE yields less in library prep, and libraries are lower molecular weight & highly duplicated compared to other sources, although % selected is similar:

Library perform & insert size are predictors of success

We also performed a partition analysis on the FFPE exome data set to identify the main statistically significant predictors of successful FFPE sequencing.

The analysis identified yield after library preparation prior to exome capture and mean insert size as main drivers in our data, and provided minimum specifications for each above which samples have the best chance for success:

Factors in model:
- PF Gb
- Library Yield
- % Duplication
- % Selected
- Fold 80 Penalty

Predicting & improving FFPE performance

Kapa Genomic QC assay for FFPE can be used as a predictor of library yield (A) and Mean Insert Size (B) to allow for triage of FFPE genomic samples before library prep.

(C) We compared different Covaris shearing conditions to assess if more gentler shearing could improve FFPE library prep success. Gentler shearing did increase yields for FFPE but did not increase insert size as it did for high quality DNA. This experiment is currently being repeated on more FFPE samples to verify reproducibility.

Conclusions

• FFPE samples routinely require more Gb to reach the same coverage levels as non-FFPE & insert size and library yield can be predictors of success for meeting target coverage deliverables.
• FFPE have a much higher rate of C>T deamination artifacts and are much more variable for other artifactual base changes.
• In lab improvements can help increase success rate and decrease artifacts.

Measurement of artifactual mutations in FFPE samples

In Q4 2015, the Picard Analysis Pipeline was updated to now calculate artifact “Q-scores” for all possible base changes in addition to the Deamination artifacts metrics it already provides (G>T/C>A).

As expected, FFPE samples have significantly higher rates of Deamination C>T/G>A artifacts (lower Q-score) & are more variable across all contexts:

DNA repair can reduce rate of deamination artifacts

We previously tested repair of FFPE using the PreCR Repair Mix (New England Biolabs), and observed modest reductions in the rate of deamination artifacts in FFPE samples we repaired:

<table>
<thead>
<tr>
<th>FFPE Sample A</th>
<th>Baseline Rate of Deamination Artifacts (Gb)</th>
<th>Rate of Deamination following PreCR (Gb)</th>
<th>Fold Reduction in Artifacts</th>
</tr>
</thead>
<tbody>
<tr>
<td>FFPE Sample A</td>
<td>1 in 1259 bases (Q31)</td>
<td>1 in 1585 bases (Q32)</td>
<td>1.26</td>
</tr>
<tr>
<td>FFPE Sample B</td>
<td>1 in 794 bases (Q29)</td>
<td>1 in 1529 bases (Q31)</td>
<td>1.58</td>
</tr>
<tr>
<td>FFPE Sample C</td>
<td>1 in 1000 bases (Q30)</td>
<td>1 in 1259 bases (Q31)</td>
<td>1.26</td>
</tr>
</tbody>
</table>

We are now following up with New England Biolabs’ newer NEBnext FFPE Repair Kit, which is more optimized for NGS library prep, to determine if we can achieve a greater reduction in the rate of these artifacts (data pending).

For FFPE, the relationship between Gb & Target Coverage is less predictable:

FFPE requires 25% more sequencing on average, & often much more!