The next generation of microsatellite instability testing

Richard Gallon PhD MBioch
Research Associate, Cancer Prevention Research Group
Institute of Genetic Medicine, Newcastle University
richard.gallon@newcastle.ac.uk
Overview and Disclosures

1. How and why we test for microsatellite instability (MSI)

2. A tumour-sequencing MSI assay, using as few as 6 markers for high throughput diagnostics

Newcastle University owns a patent covering the extended marker set used in the assay
*Patent ID: PCT/GB2017/052488, published 1st March 2018*

Named inventor on a patent filed by Newcastle University covering a reduced marker set
*PCT application number: PCT/GB2019/052148, unpublished, filing date 31st July 2019*
Clinical Needs for Mismatch Repair Deficiency Testing

Mismatch repair (MMR) deficiency testing of colorectal cancers (CRCs) is recommended to screen for Lynch syndrome (LS) (Balmana et al., 2013; Stoffel et al., 2015; UK NICE DG 27, 2017)

Screening for LS in endometrial cancers is also cost effective (Snowsill et al., 2019)

Pembrolizumab/Keytruda (immune checkpoint inhibitor) is FDA-approved as a second line treatment in any metastatic MSI-H cancer (MERCK & Co. Inc, 2017)
Tests for Mismatch Repair Deficiency

1) 4-panel IHC:

![αMSH2](image)


2) MSI testing by PCR fragment length analysis:

![Promega MSI Analysis System v1.2; Alhilal PhD Thesis, 2016](image)

3) MSI testing by next generation sequencing:

...TTGATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT TT
Tests for Mismatch Repair Deficiency

Test performance:

<table>
<thead>
<tr>
<th></th>
<th>IHC</th>
<th>MSI by FLA</th>
<th>MSI by NGS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>&gt;95%</td>
<td>&gt;95%</td>
<td>&gt;95%</td>
</tr>
<tr>
<td>Specificity</td>
<td>&gt;95%</td>
<td>&gt;95%</td>
<td>&gt;95%</td>
</tr>
<tr>
<td>Cost (inc. overheads)</td>
<td>235€</td>
<td>226€</td>
<td>607(±207)€</td>
</tr>
<tr>
<td>Throughput</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Automated</td>
<td>Not routinely</td>
<td>Not routinely</td>
<td>Yes</td>
</tr>
<tr>
<td>Other markers tested</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>LS Screening</td>
<td>Multi-step</td>
<td>Multi-step</td>
<td>One-step</td>
</tr>
</tbody>
</table>

(Shia, 2008; Zhang 2008; Kautto et al, 2016; Zhu et al, 2018; Hampel et al, 2018; Marino et al, 2018; Snowsill et al, 2019)

Only 43% of young (30-49yrs) CRC patients are screened for LS (US statistics; Shaikh et al, 2018)

...lack of testing attributed to cost (33.3%), unfamiliarity interpreting results (29.2%), and unavailable genetic counselling (24.9%) (questionnaire of 509 gastroenterologists; Noll et al, 2018)
## Considerations for MSI testing by NGS

<table>
<thead>
<tr>
<th>Sequencing platform</th>
<th>Error rate (e.g. Ion Torrent technology has high error microsatellite sequencing)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Availability (e.g. the dominance of Illumina platforms)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Marker identity</th>
<th>Sensitivity and specificity:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Length and motif (e.g. mono-, di-, tri-, tetra-nucleotide repeats)</td>
</tr>
<tr>
<td></td>
<td>• Genomic context</td>
</tr>
<tr>
<td></td>
<td>• Error rate</td>
</tr>
<tr>
<td></td>
<td>• Polymorphisms</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Classification method</th>
<th>Calling instability at individual markers:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Matched normal DNAs, or reference set of control samples</td>
</tr>
</tbody>
</table>

Classification by proportion of markers that are “unstable”:
• All markers are equal...

<table>
<thead>
<tr>
<th>Marker number</th>
<th>Variable thresholds with different marker panels</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Threshold uncertainty with small (&lt;20) marker panels</td>
</tr>
<tr>
<td></td>
<td>large panels = high cost</td>
</tr>
</tbody>
</table>
Marker selection from TCGA data:

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mononucleotide repeats</td>
<td>Greatest sensitivity and specificity (Bacher et al, 2004)</td>
</tr>
<tr>
<td></td>
<td>Detection of MSH6 deficiency (You et al, 2010)</td>
</tr>
<tr>
<td>Short (7-12bp)</td>
<td>Reduced sequencing error (Fazekas et al, 2010)</td>
</tr>
<tr>
<td></td>
<td>Greatest discrimination between MSI-H and MSS samples by NGS (Maruvka et al, 2017)</td>
</tr>
<tr>
<td>Common SNP within 30bp</td>
<td>To assess allelic distribution of microsatellite deletions</td>
</tr>
<tr>
<td>Monomorphic</td>
<td>No need for matched normal DNA</td>
</tr>
</tbody>
</table>

120 candidate markers identified.
A Naïve Bayesian Approach to MSI Classification

Library preparation:

- **PCR$_1$:** Singleplex amplification
  - Amplicon pooling per sample
  - Amplicon purification (AMPure XP Beads, Beckman Coulter)
  - **PCR$_2$:** Addition of sample indexes and sequencing adaptors
  - Amplicon purification, dilution and pooling

Sequencing and analysis:

- Amplicon sequencing, target depth of 5000x (MiSeq, Illumina)
  - Read alignment to reference genome (BWA, Li & Durbin, 2010)
    - Detection of microsatellite deletion frequency and allelic distribution

Discovery cohort:

- 58 CRCs of known MSI status*
- Selection of 17 markers based on ROC AUCs

*MSI by FLA (Promega)
A Naïve Bayesian Approach to MSI Classification

Classifier training cohort:

• 139 CRCs of known MSI status*

• MSI classification using microsatellite deletion frequency and allelic distribution

*MSI by FLA (Promega)
A Naïve Bayesian Approach to MSI Classification

Classifier training cohort:

- 139 CRCs of known MSI status*
- MSI classification using microsatellite deletion frequency and allelic distribution

*MSI by FLA (Promega)
A Naïve Bayesian Approach to MSI Classification

Classifier training cohort:

- 139 CRCs of known MSI status*
- MSI classification using microsatellite deletion frequency and allelic distribution
- 100% sensitivity and 100% specificity

Validation cohort:

- 70 CRCs of known MSI status*
- 97% sensitivity and 97% specificity
- Discordance resolved in 3 of 4 samples

*MSI by FLA (Promega)

Probability from assay observations:

\[
p(O|MSI) \frac{p(O|MSI)}{p(O|MSS)} = \prod_{i=1}^{N} p(O_i|MSI) \frac{p(O_i|MSI)}{p(O_i|MSS)}
\]

Calculating assay score:

\[
score = \log_{10} \left( \frac{p(\text{MSI}) \cdot p(O|MSI)}{p(\text{MSS}) \cdot p(O|MSS)} \right)
\]

prior probability = 0.15/0.85
A Low Cost Sequencing-based MSI Assay for Clinical Diagnostics

24 MNRs and *BRAF* multiplexed by single molecule molecular inversion probes (smMIPs):

Followed recommendations for the validation of NGS-based oncology assays for clinical diagnostics (Association of Molecular Pathology, College of American Pathologists; Jennings *et al*, 2017).
A Low Cost Sequencing-based MSI Assay for Clinical Diagnostics

Validation of diagnostic accuracy:

- Sample number and type
- Independent cohorts
- Reproducibility

Quality controls (QCs):

- Detection limits
- Internal quality checks of library complexity

Clinical utility:

- Sample identification
A Low Cost Sequencing-based MSI Assay for Clinical Diagnostics

Classifier training and validation:

100% sensitivity and 100% specificity using either 24 or 6 markers
A Low Cost Sequencing-based MSI Assay for Clinical Diagnostics

Assessing detection limits using sample mixtures:

3-6% MSI-H cell line DNA can be detected in mixtures – equivalent to MSI by FLA
A Low Cost Sequencing-based MSI Assay for Clinical Diagnostics

Molecular barcodes (MBs) provide an internal QC of sequencing:

>75 MBs/marker ensures results reliability
A Low Cost Sequencing-based MSI Assay for Clinical Diagnostics

Assessing assay reproducibility and portability:

Research laboratory

- 32 CRCs repeat tested from amplification through to classification

<table>
<thead>
<tr>
<th>24 marker panel</th>
<th>6 marker panel</th>
</tr>
</thead>
<tbody>
<tr>
<td>100% concordance</td>
<td>100% concordance</td>
</tr>
<tr>
<td>Score correlation $R^2 = 0.97$</td>
<td>Score correlation $R^2 = 0.97$</td>
</tr>
</tbody>
</table>

3 samples had <75 MBs/marker but still correctly classified, in-line with in silico predictions.
A Low Cost Sequencing-based MSI Assay for Clinical Diagnostics

Assay summary:

<table>
<thead>
<tr>
<th></th>
<th>24 MNRs plus Braf</th>
<th>6 MNRs plus Braf</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Accuracy</strong></td>
<td>&gt;95% sens. + spec.</td>
<td>&gt;95% sens. + spec.</td>
</tr>
<tr>
<td><strong>Detection limit</strong></td>
<td>&gt;3% MSI-H DNA</td>
<td>&gt;6% MSI-H DNA</td>
</tr>
<tr>
<td><strong>Quality control</strong></td>
<td>&gt;75 MBs/marker</td>
<td>&gt;75 MBs/marker</td>
</tr>
<tr>
<td><strong>LS screening</strong></td>
<td>One-step</td>
<td>One-step</td>
</tr>
<tr>
<td><strong>Sample identification</strong></td>
<td>pr(SNP match) = 3.6x10^{-10}</td>
<td>pr(SNP match) = 3.8x10^{-3}</td>
</tr>
<tr>
<td><strong>Reagent cost</strong></td>
<td>&lt;20€ per sample</td>
<td>&lt;10€ per sample</td>
</tr>
</tbody>
</table>

Future work:

- Commercialisation (Cancer Research UK Commercial Partnerships)
- Continued validation and accreditation (Northern Genetics Service, Newcastle Hospitals)
- Application to extra-colonic cancers
### Tumour Type

<table>
<thead>
<tr>
<th>Tumour Type</th>
<th>Percentage MSI-H (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast</td>
<td>50.0% (21.5% - 78.4%)</td>
</tr>
<tr>
<td>Colorectal</td>
<td>94.0% (83.8% - 97.9%)</td>
</tr>
<tr>
<td>Endometrial</td>
<td>74.4% (58.9% - 85.4%)</td>
</tr>
<tr>
<td>Other GI</td>
<td>50.0% (2.6% - 97.4%)</td>
</tr>
<tr>
<td>Ovarian</td>
<td>85.7% (48.7% - 99.3%)</td>
</tr>
<tr>
<td>Prostate</td>
<td>0.0% (0.0% - 94.9%)</td>
</tr>
<tr>
<td>Sebaceous Adenoma</td>
<td>100.0% (67.6% - 100.0%)</td>
</tr>
<tr>
<td>Other Skin</td>
<td>40.9% (38.7% - 76.7%)</td>
</tr>
<tr>
<td>Urothelial</td>
<td>100.0% (56.6% - 100.0%)</td>
</tr>
</tbody>
</table>
Detection of Constitutional Mismatch Repair Deficiency

Low-level MSI in non-neoplastic tissues is detectable with an alternative analysis pipeline:

ROC AUC: 1.00

Dotted lines represent *a priori* score thresholds of 1.30 (95% probability not a control sample), and 2.00 (99% probability not a control sample).

Using a threshold of 2.00, the assay has 97% sensitivity, and 100% specificity.

Notes:

† homozygous for a hypomorphic PMS2 variant

§ patient 8 (3 samples all with low scores) was recovering from aplasia at sample collection
What is the next generation of MSI testing?

Heterogeneous, due to:

• New developments – e.g. automation of IHC
• Compatibility with established clinical pathways
• Cost versus resource
• Local expertise
• Local infrastructure

Diagnostic niche for low cost sequencing-based MSI assays


richard.gallon@newcastle.ac.uk
Special thanks to:

The Barbour Foundation for funding my PhD.

My PhD supervisory team, Mauro Santibanez-Koref, Mike Jackson, and John Burn.

My colleagues Christine Hayes, and Harsh Sheth.

richard.gallon@newcastle.ac.uk
Additional thanks to:

The CaPP3 clinical trial team: Gill Borthwick, Donna Job, Jackie Greenwood, Lynn Reed, Lynne Longstaff, and Amy McAllister.


richard.gallon@newcastle.ac.uk