Accurate genotyping of UGT1A1 dinucleotide repeat polymorphism from targeted NGS data for the assessment of irinotecan chemotherapy adverse events

Francisco M. De La Vega1, Len Trigg2, Kurt Gaaster2, Sean A. Irvine2, Gene Selkova1, Yan Yang1, Kyung Choi1, and Robert Huether1

1Tempus Labs, Chicago, IL, USA; 2Real Time Genomics, Ltd., Hamilton, New Zealand

INTRODUCTION

Colorectal cancer (CRC) is a leading cause of cancer-related death across the world. Irinotecan (IRI) is commonly used to treat metastatic CRC. The gene UGT1A1 encodes the enzyme responsible for the glucuronidation of SN-38, the active metabolite of IRI. The TA repeat in the promoter region of UGT1A1 is highly polymorphic. Wild-type UGT1A1 contains six TA repeats (TA6(TA)4). Polymorphic UGT1A1 alleles with a higher number of TA repeats, such as UGT1A1*28 (TA14(TA)8) and *17(TA)4 alleles, decrease promoter activity and are associated with severe toxicity in patients receiving IRI-based chemotherapy, for which dose reductions are recommended. Matched tumor/normal genomic profiling by NGS for cancer therapy may be useful to assess therapy-induced adverse events due to germline variants such as those in UGT1A1. However, genotyping of UGT1A1 polymorphisms is commonly carried out with PCR or fragment analysis in capillary electrophoresis, and not from NGS data. This is due to challenges in aligning short reads to repeats and the introduction of “stutter” artifacts due to DNA polymerase slippage that add or delete a single nucleotide during PCR. Inaccurate stutter models and stutter-free algorithms are used to align short reads to repeats and identify UGT1A1 repeat genotypes from target capture NGS data and demonstrate the feasibility of this method for genomics profiling of cancer patients.

METHODS

BayeSTR analyzes deduplicated read alignments to a graph-based model representing the possible repeat alleles2, and then performs genotype calling by a Bayesian model that incorporates an empirically derived DNA polymerase stutter densimizing model. The Bayesian model provides genotype posterior probabilities as confidence values that can be used to eliminate genotyping errors for poor quality data/samples.

RESULTS

Figure 3. Stutter spectrum error model

The empirically observed stutter distribution is well approximated by a simple one-parameter (probability of T insertion, σ) model. The model can generate arbitrary repeat-lengths for which empirical data was not available. This model is used to construct priors when evaluating genotype hypothesis by BayeSTR (see Table 2).

Table 2. Performance with simulated data

<table>
<thead>
<tr>
<th>Gene</th>
<th>Repeat</th>
<th>Genotype</th>
<th>Cov</th>
<th>Call</th>
<th>Match</th>
<th>Match Mismatch</th>
</tr>
</thead>
<tbody>
<tr>
<td>UGT1A1</td>
<td>70X</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>UGT1A1</td>
<td>50X</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>UGT1A1</td>
<td>30X</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>UGT1A1</td>
<td>10X</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Validation

We sequenced SF reference Coriell cell lines previously characterized by the CDC GHRM project with orthogonally validated UGT1A1 repeat alleles. We compared calls by BayeSTR with those made by the Expansion Hunter2 software. BayeSTR calls matched the truth set, except for NA06986, where a 588 was also present in the repeat. By contrast, Expansion Hunter exhibits a significantly lower rate for 1,428/298 homogametes. Data from cell lines with 588 and 577 genotypes were not evaluated with Expansion Hunter.

CONCLUSIONS

BayeSTR allows for automated and accurate UGT1A1 promoter genotyping from targeted NGS data and can be applied to other genomic repeat regions of clinical relevance.

This method identifies UGT1A1 repeat polymorphisms associated with IRI-induced adverse events and can be used during clinical NGS testing to further support clinician treatment decisions for cancer patients.

REFERENCES