SureSeq

The application of a one-day hybridisation-based enrichment protocol for next-generation sequencing incorporating a rapid (30 minute) hybridisation step

Lyudmila Georgieva, Ezam Uddin, Jacqueline Chan, Faidra Partheniou and Graham Speight [1]Oxford Gene Technology (OGT).

A Sysmex Group Company

What binds us, makes us.

Introduction

- Hybridisation-based enrichment protocols for next-generation sequencing (NGS) generate higher quality data (e.g. enhanced coverage uniformity, more complete coverage, and more accurate assessment of insertions/deletions (indels) and internal tandem duplications (ITDs)). However, they are generally more time consuming than PCR-based enrichment approaches.
- We have developed a rapid (30 minute) hybridisation protocol that enables Illumina sequencer-ready libraries to be generated from purified DNA in 1-day.
- The aim of this study is to evaluate the new streamlined 1-day hybridisation-based NGS library preparation kit (LPK) in conjunction with four different haematological capture panels of varying sizes.

Methods

Preparation of purified DNA to sequencer-ready libraries in 7 hours, 45 minutes

- An enhanced version of the SureSeq™ LPK (OGT) was utilised which incorporates an enzymatic DNA fragmentation in combination with a rapid hybridisation of just 30 minutes. This enhanced protocol reduces the overall processing time by 6 hours, resulting in a streamlined, 1-day workflow.
- This kit offers a similar turn-around time to amplicon-based enrichment protocols, without the associated disadvantages, such as PCR bias, allelic bias (indels) and drop-outs, as well as poor uniformity of coverage.

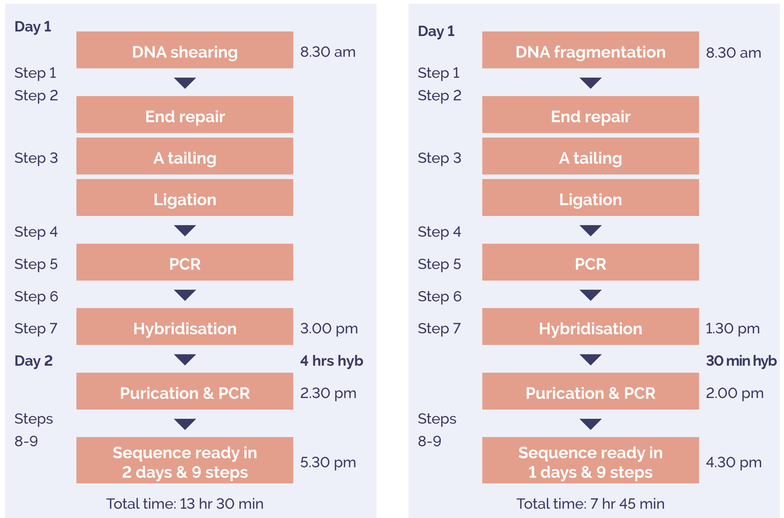

Figure 1: Comparison of workflows.

Study design

- Four different haematological panels have been used, with a size range from 0.5 Kb to 138 Kb.
- Data quality comparison was performed in terms of Mean Target Coverage (MTC) and % on-target bases achieved with the 1-day and the standard workflow. More specifically, we compared the uniformity of coverage achieved with both protocols for difficult to sequence genes such as *CALR*, *CEBPA* and *FLT3*, as well as the coverage at the key myeloproliferative neoplasm mutation sites: *JAK2 V617F*, *JAK2* exon 12, *MPL* W515K/L and *CALR* exon 9.
- Sequencing was performed on a MiSeq® using a V2 300 bp cartridge (Illumina).

#	Panel name	Target size (Kb)
1	SureSeq Core MPN	0.5
2	SureSeq Myeloid	41
3	SureSeq myPanel™ Acute Myeloid Leukaemia (AML)	52
4	SureSeq myPanel™ HaemOnc	138

Table 1: Panel names and sizes.

Results

Comparison of the data generated by the 1-day and standard NGS protocols

- Data presented here are from 24* samples that were processed using the enhanced LPK in combination with four haematological panels on an Illumina MiSeq.
- The quality of the data generated with the 1-day protocol is comparable to the standard 4-hour hybridisation protocol.
- OGT 1-day protocol generated >85% of the % on-target bases generated with the standard protocol. The % change is consistent for all panel sizes.

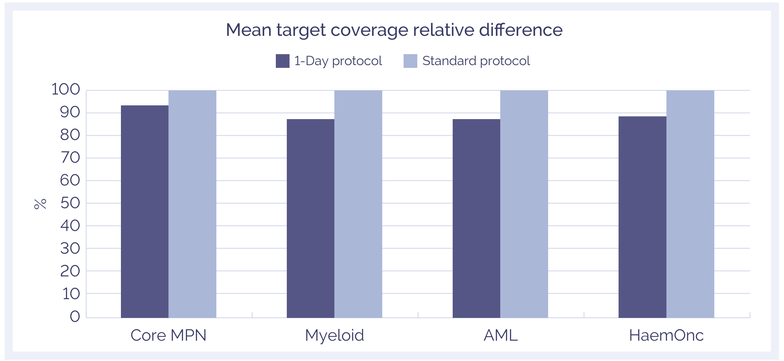

Figure 2: On-target rate comparison between 1-day and standard NGS protocol.

The MTC generated is dependent on the size of each panel. Overall, both workflows generated very good coverage. The MTC generated with the 1-day protocol is >80% of the MTC generated with the standard protocol. The % change is consistent for all panels.

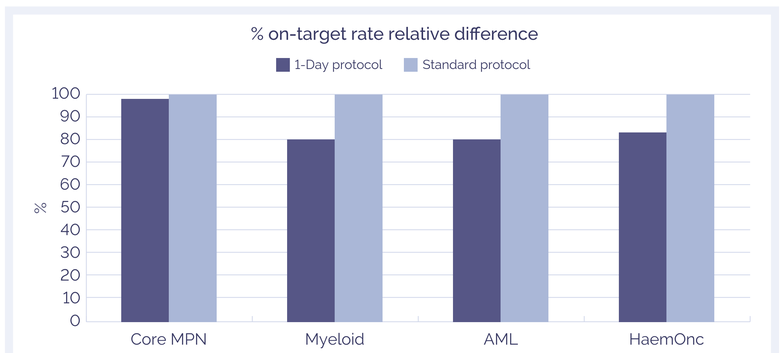

Figure 3: Mean target coverage comparison between 1-day and standard NGS protocol.

All panels meet the following uniformity specifications: >99% of bases covered at >20% of the mean (after de-duplication). This permits the reliable detection of more complex rearrangements (i.e.) indels and ITDs.

Accurate and reproducible variant detection even in heterogeneous samples

The SureSeq Core MPN panel has been validated with samples from the National Institute for Biological Standards and Control (NIBSC) and we have shown the accurate detection of JAK2 V617F is possible down to the 1% Variant Allele Frequency (VAF) level at a de-duplicated read depth of >1000x (Table 2).

NIBSC *JAK2* V617F Sample	% VAF	Wild type reads	Variant reads
10%	13.1	1038	157
5%	4.6	1623	78
1%	1.0	2172	23
0.1%	0.4	2414	9
0%	0.0	2555	1

Table 2: Data generated from a 48 sample run on an Illumina MiSeq. The SureSeq Core MPN panel in conjunction with the 1-day protocol permitted the detection of alleles at 1% VAF with high confidence.

Accurate detection of difficult to sequence genes

Mutations in the *CEBPA* and *FLT3* genes are among the most common molecular alterations in AML. Sequencing of the *CEBPA* gene is often hampered by a repetitive nucleotide sequence and a very high GC-rich content. Genes such as *FLT3* ITDs are challenging to target because they are by nature repetitive, can be long and are generally masked in most panel designs.

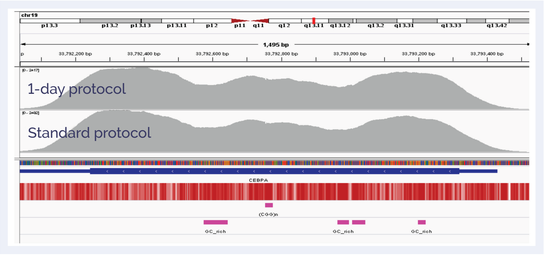

Figure 4: Excellent uniformity of coverage of the *CEBPA* gene averaging ~2000x coverage. Depth of coverage per base (grey). GC percentage (red). Repeat regions and GC-rich regions (pink). Data shown from 1-day protocol.

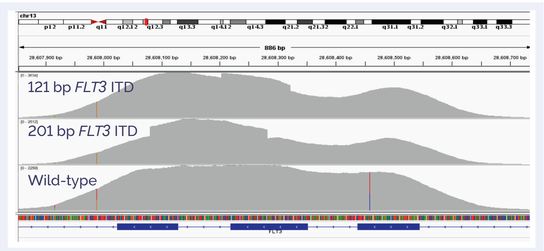

Figure 5: Detection of 121 bp and 201 bp *FLT3* ITD. Wild-type sample (bottom panel).

Accurate detection of deletions

Using the enhanced workflow we were able to reliably detect single nucleotide variants (SNVs) as well as insertions (5 bp insertion in *JAK2* exon 12 and *CALR* exon 9) and deletions (5 bp deletion exon 12 *JAK2* and 52 bp deletion *CALR* exon 9).

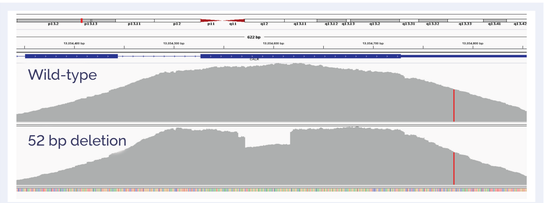

Figure 6: Detection of a 52 bp deletion (exon 9 *CALR*). Wildtype sample (top panel) is compared to a 52 bp somatic deletion (bottom panel). Data shown from 1-day protocol.

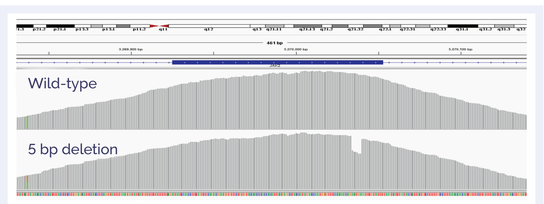

Figure 7: Detection of a 5 bp deletion (exon 12 *JAK2*). Wild-type sample (top panel) is compared to a 5 bp somatic deletion (bottom panel). Data shown from 1-day protocol.

Conclusions

- We have successfully utilised the OGT 1-day hybridisation-based SureSeq LPK protocol in combination with four haematological cancer panels to reliably and routinely detect somatic SNVs by NGS down to a 1% VAF.
- The uniformity of coverage of this approach permitted the detection of key *CALR* and *JAK2* indels (including 52 bp deletions and 5 bp insertions) and *FLT3* ITDs to be identified.
- This enhanced protocol incorporates an enzymatic fragmentation step which permits the highthroughput preparation of 24-48 samples (panel size dependent) from genomic DNA to sequencer in a 1-day workflow.
- To achieve >1000x de-duplicated depth (required for confident detection of 1% VAF), 24-48 samples (panel size dependent) can be reliably sequenced in a single MiSeq (V2 300 bp) run. This allows the generation of high quality data in a cost effective and timely manner.

Acknowledgements *Samples kindly provided by Prof. Nick Cross (National Genetics Reference Laboratories - Wessex, UK)

Further information **contact@ogt.com | ogt.com | Oxford Gene Technology Ltd., Begbroke Science Park, Woodstock Road, Begbroke, Oxfordshire, OX5 1PF, UK. | SureSeq**: For Research Use Only; Not for Diagnostic Procedures.

990301 01/21