SureSeq

The accurate detection by next-generation sequencing (NGS) of difficult to sequence genes (*CALR, CEBPA, FLT3*) associated with myeloid disorders using a hybridisation-based enrichment approach

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

A Sysmex Group Company

What binds us, makes us.

Introduction

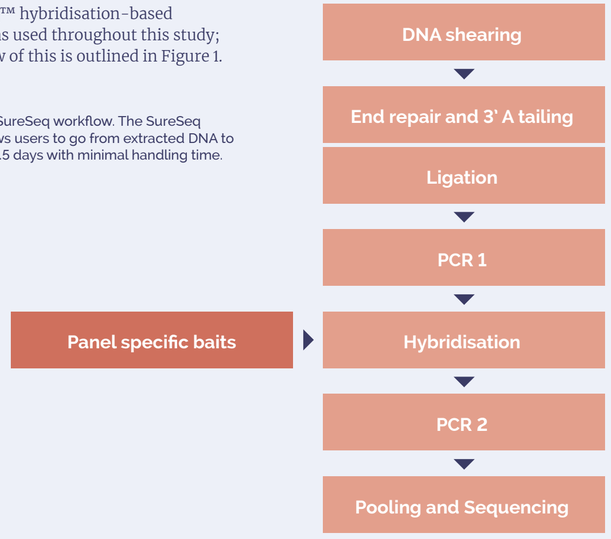
It has been reported that mutations in the CCAAT/enhancer binding protein alpha (*CEBPA*) and fms-related tyrosine kinase 3 (*FLT3*) genes are among the most common molecular alterations in acute myeloid leukaemia (AML). Two mutations on separate alleles of *CEBPA* with a specific combination of an N-terminal frameshift mutation on one allele and a C-terminal in frame mutation on the other allele are important for prognosis[2,3]. The prevalence of an internal tandem duplication (ITD) of the juxtamembrane domain-coding sequence and a missense mutation of D835 within the kinase domain of the *FLT3* gene occurs in 15-35% and 5-10% of adults with AML, respectively[4]. A quarter of patients with essential thrombocythemia or primary myelofibrosis carry a driver mutation of calreticulin (*CALR*). A 52-bp deletion (type 1) and a 5-bp insertion (type 2 mutation) are the most frequent variants[5].

Further research to better characterise the involvement of these genes in acute myeloid leukaemia (AML) would be beneficial. NGS is one technology being explored for further research into myeloid disorders such as myeloproliferative neoplasms (MPNs) and AML but has been hampered by the difficulty in sequencing certain genes. Amongst the difficult to sequence genes are *CALR*, *CEBPA* and *FLT3*. The development of robust assays for *CEBPA* mutation analysis is challenging due to the high GC content of the gene (75% in the coding region), the presence of a trinucleotide repeat region, the complexity of the mutations, and the frequent occurrence of mutations in mononucleotide repeats. Genes such as *FLT3* are challenging to target because they contain complex repetitive elements that can be long and are generally masked in most panel designs. *CALR* sequencing is challenging due to the presence of low complexity regions making the detection of insertions and deletions difficult.

Methods

The SureSeq™ hybridisation-based approach was used throughout this study; the workflow of this is outlined in Figure 1.

Figure 1: OGT SureSeq workflow. The SureSeq workflow allows users to go from extracted DNA to sequencer in 1.5 days with minimal handling time.



Panel content

We have utilised a hybridisation-based enrichment approach in combination with a SureSeq myPanel™ NGS Custom AML Panel

Study Design

The panel was used to confirm variants in research samples* containing variants in each of these difficult to sequence genes. Sequencing was conducted on a MiSeq® using a V2 300 bp cartridge (Illumina).

Hybridisation-based enrichment generates highly uniform coverage of key targets

- Heterogeneous cancer samples pose significant challenges as alleles are likely to be present at a lower fraction than what would be expected for germline variants. Samples typically contain a mixture of cancer and normal cells; moreover, the tumour may consist of several molecularly distinct clones. In order to detect alleles that contribute only a small percentage to the reads at any locus, a highly uniform and sensitive enrichment is required.
- Uniformity of coverage is a useful value with which to compare this distribution and can be expressed as the percentage of target bases that have greater than 20% of the mean coverage.
- The SureSeq myPanel NGS Custom AML Panel in our hands met the following uniformity specifications: >99% of bases covered at >20% of the mean (after removal of PCR duplicates). This permits the reliable detection of more complex rearrangements (i.e.) indels and ITDs.

Accurate detection of variants in difficult to sequence genes

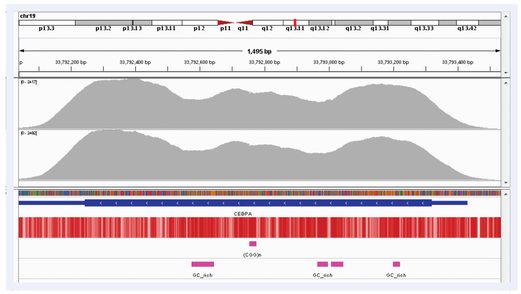


Figure 2: Excellent uniformity of coverage of the *CEBPA* gene (average depth~2000x). Depth of coverage per base (grey). GC percentage (red). Repeat regions and GC-rich regions (pink).

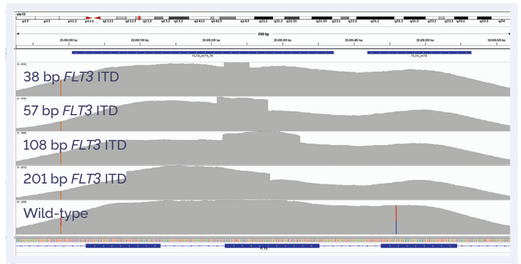


Figure 3: Detection of *FLT3*-ITDs with different sizes. Wild-type sample (bottom panel).

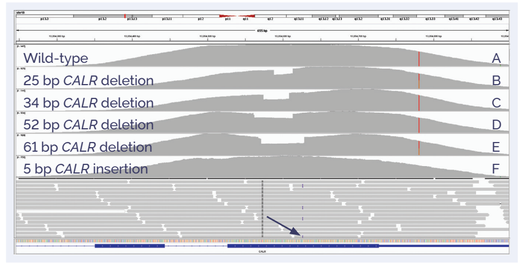


Figure 4: Detection of insertions and deletions in exon 9 of *CALR*. Wild-type sample (panel A) is compared to a 25 bp (panel B), 34 bp (panel C), 52 bp (panel D), 61 bp somatic deletions (panel E) and 5 bp insertion (panel F).

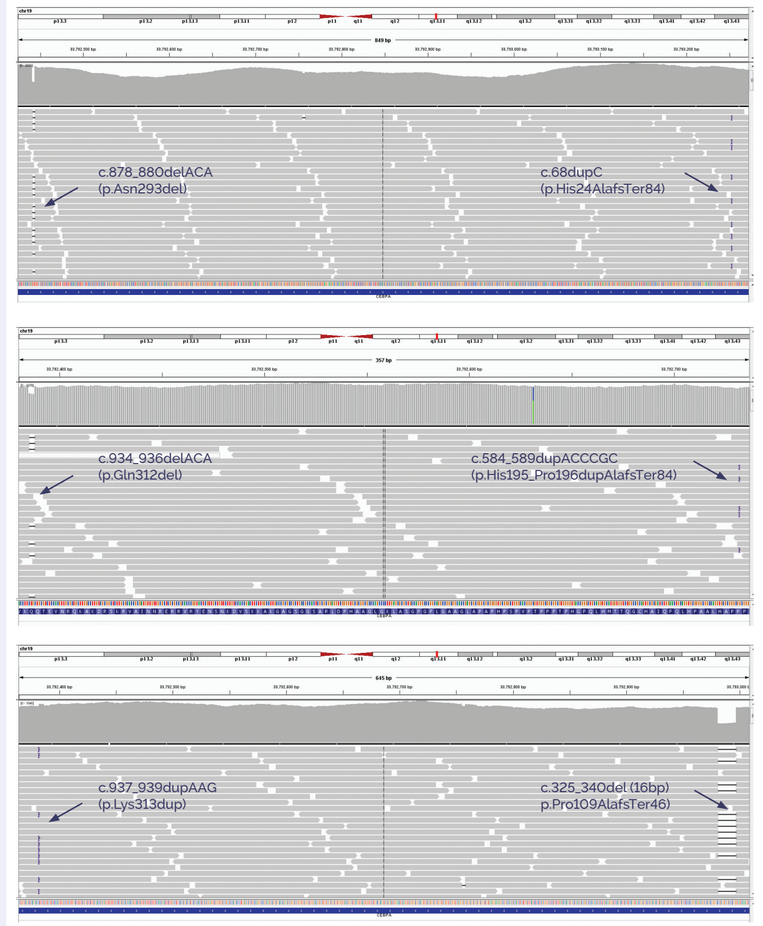


Figure 5: Detection of insertions and deletions in *CEBPA*.

Conclusions

- Excellent uniformity of coverage was obtained from the hybridisation-based enrichment using a SureSeq myPanel NGS Custom AML Panel.
- This approach is a robust method that can be used successfully in sequencing difficult genes (including *CEBPA*, *FLT3* and *CALR*).
- High levels of uniformity are maintained across all genes permitting the reliable detection and accurate sizing (including low allele frequency) of key *CALR* variants (including 52 bp deletions and 5 bp insertions), SNVs, indels and other in frame variants throughout *CEBPA* with a de-duplicated depth of greater than 2000x as well as ITDs of between 24 and 201 bp in *FLT3*.
- This approach may therefore remove the requirements for supplementary approaches to analyse these difficult genes, such as Sanger sequencing *(CEBPA)* and fragment analysis *(CALR and FLT3)*.

References

2. Behdad A, *et al.*, A clinical grade sequencing-based assay for *CEBPA* mutation testing: report of a large series of myeloid neoplasms. *J Mol Diagn.* 2015 Jan;17(1):76-84. | 3. Spencer DH, *et al.*, Detection of *FLT3* internal tandem duplication in targeted, short-read-length, next-generation sequencing data. *J Mol Diagn.* 2013 Jan;15(1):81-93. | 4. Dufour A, *et al.*, Acute myeloid leukemia with biallelic *CEBPA* gene mutations and normal karyotype represents a distinct genetic entity associated with a favorable clinical outcome. *J Clin Oncol.* 2010 Feb 1;28(4):570-7. | 5. Pietra D, *et al.*, Differential clinical effects of different mutation subtypes in *CALR*-mutant myeloproliferative neoplasms. *Leukemia.* 2016 Feb; 30(2): 431–438.

Acknowledgements *Research samples provided by the National Genetics Reference Laboratories - Wessex, University of Virginia, University of Newcastle and UK NEQAS.

Further information **contact@ogt.com | ogt.com | Oxford Gene Technology Ltd., Begbroke Science Park, Woodstock Road, Begbroke, Oxfordshire, OX5 1PF, UK. | SureSeq**:

990302 01/21