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 and Graham Speight
Oxford Gene Technology (OGT), Begbroke Science Park, Begbroke Hill, Woodstock Road, Begbroke, Oxford, UK

Introduction
The application of short read NGS for research into myeloid disorders such as myeloproliferative neoplasms (MPNs) and acute myeloid leukaemia (AML) has been hampered by the difficulty in sequencing certain genes. Amongst the difficult to sequence genes are calreticulin (CALR), CCAAT/enhancer binding protein alpha (CEBPA) and fms-related tyrosine kinase 3 (FLT3).
The development of robust assays for CALR mutation testing is challenging due to the high GC content of the gene (75% in the coding region), the presence of a thymidylate repeat region, the complexity of the mutations, and the frequent occurrence of mutations in mono nucleotide 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.

Mutations in the CEBPA and FLT3 genes are among the most common molecular alterations in 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(1). 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(2). A quarter of patients with essential thrombocythemia or primary myelofibrosis carry a driver mutation of CALR. A 52 bp deletion (type 1) and a 5 bp insertion (type 2 mutation) are the most frequent variants(3).

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

Conclusions
• Excellent uniformity of coverage was obtained from the hybridisation-based enrichment using the 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), SNS, 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

Acknowledgements
*Research sample provided by the National Genetics Reference Laboratories – Wessex, University of Virginia, University of Newcastle and UK NEQAS