Reference Materials for Measurable Residual Disease (MRD) Monitoring

Yves Konigshofer, Matthew G. Butler, Jessica Dickens, Omoshile Clement, Andrew T. Anfora, Dan Brudzewsky, Bharathi Anekella, and Russell K. Garlick
LGC SeraCare Life Sciences, Gaithersburg, MD

AACR 2020
June 22 – 24
Virtual Annual Meeting II

Poster # 1978
Clinical Research, Circulating Markers 2
Contact email: yves.konigshofer@lgcgroup.com

INTRODUCTION

The analytical validation of liquid biopsy-based assays that attempt to monitor for the disappearance and reemergence of cancer can be challenging due to the need for reference materials that allow for the assessment of sensitivity and specificity at variant allele frequencies (VAFs) that can be over an order of magnitude below those that can be detected reliably by typical circulating tumor DNA (ctDNA) assays. At such low VAFs, a given plasma sample may only contain limited somatic variants - and only at a few copies - which is why some MRD assays are now focusing on monitoring patient-specific somatic variants throughout the genome using patient-specific assays. This also means that validation should include the steps needed to monitor patient-specific somatic variants.

Therefore, we designed our reference materials for a tumor/normal workflow as a set of three components. First, one cell line is a source of normal DNA that can be used to assess specificity. Second, a germline SNP-matched cell line provides hundreds to thousands of additional somatic variants. Third, blends of fragmented and sized DNA are used to mimic circulating cell-free DNA (ccfDNA) and serve as the input for MRD assays at VAFs from 1% to 0.001% (one mutant copy per 0.35 to 350 ng of ccfDNA) and at 0 %.

Here, we describe the initial characterization of these reference materials.

MATERIALS AND METHODS

Cell Lines

DNA from SNP-matched cell lines were used in a tumor/normal workflow where somatic variants were present at ~heterozygous frequencies without signs of aneuploidy. To identify somatic variants, Whole Exome Sequencing (WES) was performed using SureSelect Human All Exon v6+COSMIC kits (Agilent Technologies, Santa Clara, CA). Sequencing was performed on a NovaSeq® (Illumina, San Diego, CA) at Psomagen (Rockville, MD). Resulting FASTQ files were analyzed further. DNA from the GM24385 lymphoblastoid cell line was used as a normal.

Reference Materials

DNA from SNP-matched cell lines was quantified by digital PCR and blended to target VAFs of 1%, 0.1%, 0.01%, and 0.001% for somatic variants (tumor DNA content being ~twice those VAFs). DNA was fragmented as part of the blending process to approximate the average size of ccfDNA.

Custom Assay Design

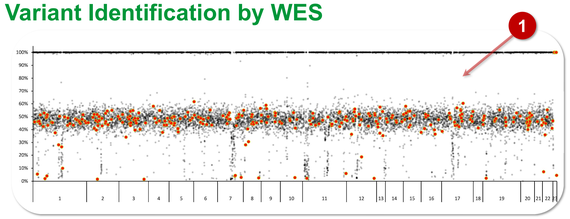
Based on past experiments and data presented at AACR's 2018 annual meeting, assays focused on A>C (T>G) and C>G (G>C) SNVs that typically had the lowest background in MiSeq® (Illumina) data. 58 such changes across the three SNP-matched tumor/normal lines were used as targets for custom SureSelect® XT HS (Agilent Technologies) and LiquidPlex™ (ArcherDX, Boulder, CO) assays. Approximately 30 ng of DNA input (~8,600 haploid genomes) was used for the NGS assays. Initial sequencing was done on a MiSeq. For the SureSelect XT HS assay, an insufficient number of reads were obtained to make use of MBCs when 8 samples were analyzed per standard flow cell, and the samples were also sequenced on a HiSeq® (Illumina) to obtain ~30x higher depth.

In order to evaluate the somatic variants and any background in SureSelect assays, paired reads were first compared in order to remove adapters and discordant bases. MBC-based error correction was then carried out using the LocatIt tool in AGeNT (Agilent). In addition to the sample data, synthetic tumor datasets were generated where all targeted SNVs were present at 100% or where mutations to A, C, G, or T were present. BAM files from the synthetic datasets were used in conjunction with the SureSelect BAM data to generate tumor/normal mpileup files with samtools (Genome Research Limited, London, UK) requiring base and mapping Q scores of 20. No BAQ adjustment was performed. The mpileup files were then analyzed with VarScan® (Koboldt, DC, et al., Genome Research, 2012, 22(3), 568-576) where the synthetic data represented the tumor and a given tumor line (or WT) represented the normal in order to ensure that the 58 possible mutation-containing sites would get reported. The resulting VCF files were then used to assess the targeted SNVs and any additional background.

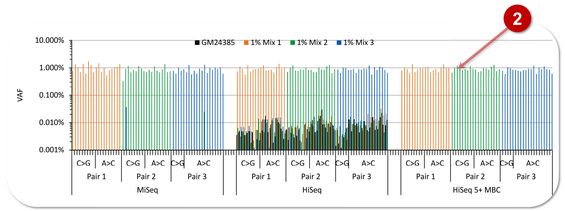
In order to evaluate the somatic variants and any background in LiquidPlex assays, custom Targeted Mutation Files were used in Archer Analysis v6 that looked for all possible SNVs at the 58 locations (e.g., A>C, A>G, and A>T). Additionally, a similar custom pipeline to the one used for SureSelect samples was used to assess the effects of MBC-based error correction.

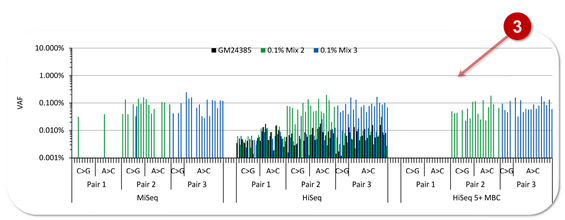
RESULTS

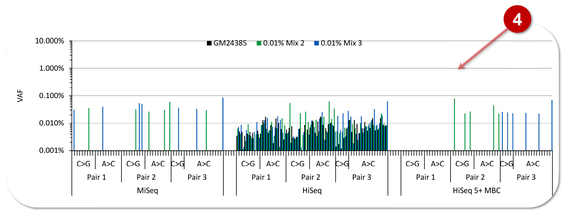
Variant Identification by WES



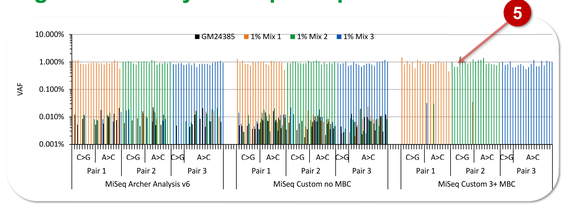
Targeted SNVs by 8 Sample SureSelect XT HS



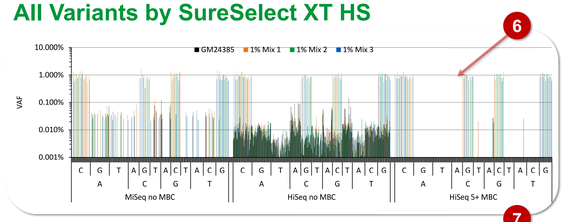


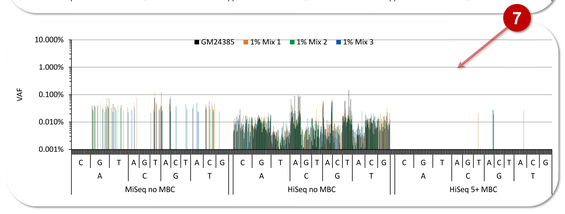


Targeted SNVs by 4 Sample LiquidPlex

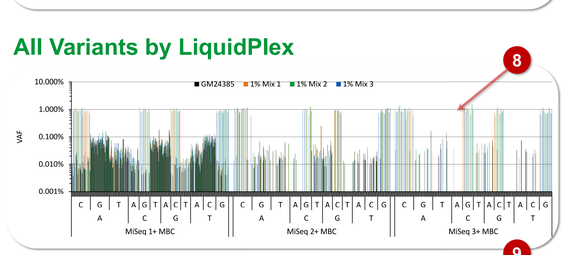


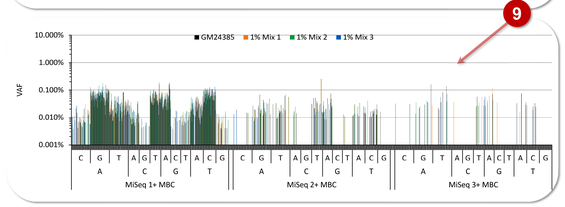
All Variants by SureSelect XT HS



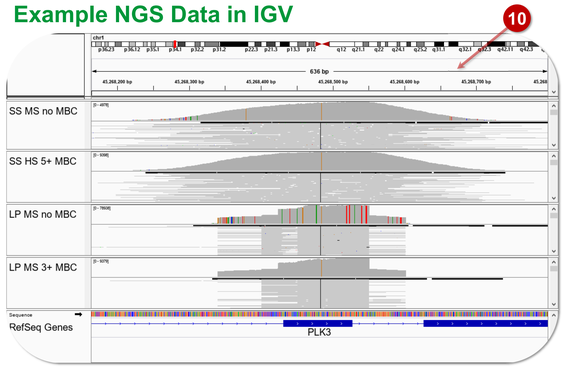


All Variants by LiquidPlex





Example NGS Data in IGV



DISCUSSION

1. SureSelect Whole Exome v6+COSMIC data is shown for one SNP-matched tumor/normal pair. Nonsynonymous SNVs with VAFs at and above 1% that were identified by VarScan as somatic are highlighted in orange. Looking for patient-specific mutations for MRD detection may start with a similar approach.
2. Data from a custom SureSelect XT HS panel is shown that simultaneously targets 58 variants across the three tumor/normal pairs. Sequencing was done on both a MiSeq and a HiSeq, where HiSeq data was also analyzed after error correction using molecules with at least 5 observations. VAFs for heterozygous somatic mutations in each of three 2% tumor content mixes were observed close to 1%. Additional observations of these variants – for example, in a different line or the GM24385 reference line – had VAFs close to 0.01% in raw HiSeq data and were frequently not observed in lower depth MiSeq data or error corrected HiSeq data.
3. Similarly, analysis of two 0.2% tumor content mixes observed most of the expected somatic mutations around 0.1% VAF, with additional background observations generally well below 0.1% VAF.
4. Analyses of two 0.02% tumor content mixes observed several expected somatic mutations above background, but background in the absence of error correction was around the 0.01% level. At 100% library incorporation efficiency, approximately 11 variants per mix may have been present at 1+ copies. At 30% library incorporation efficiency, closer to 5 variants, which is what was observed in error corrected HiSeq data.
5. Results from a custom LiquidPlex panel are shown that simultaneously target the same 58 variants. Data was analyzed in Archer Analysis and using a custom pipeline similar to that used for the SureSelect data with either no error correction or using molecules with at least 3 observations. VAFs for heterozygous somatic mutations in each of three 2% tumor content mixes were observed close to 1%. Additional observations of these variants – for example, in a different tumor line or the GM24385 reference line – were around 0.01% VAF. Such variants in error corrected data were single copy.
6. For the custom SureSelect XT HS assay, VAFs of all possible variants at the targeted locations were determined for the GM24385 and 2% tumor content mixes with and without error correction using molecules with at least 5 observations.
7. Masking of the expected 1% variants shows that the background for the A>C and C>G SNVs appears to be lower than for other SNVs on the MiSeq. A>T SNVs also had a relatively low background on the HiSeq with patterned flow cells suggesting that the sequencing process may elevate A>T noise on a MiSeq with v2 reagents. Error correction removed most noise, with remaining variants mostly C>T.
8. For the custom LiquidPlex assay, VAFs for all possible variants at the targeted locations were determined for the GM24385 and 2% tumor content mixes.
9. Masking the expected 1% variants shows that the background for the A>C and C>G SNVs appears to be around 0.01% and lower than for other SNVs. A higher background – closer to 0.1% VAF for some variants – was observed than with the SureSelect XT HS assay when including uncorrected reads.
10. NGS data is shown in IGV for SureSelect (SS) and LiquidPlex (LP) MiSeq (MS) and HiSeq (HS) platforms with and without MBC error correction for the same 2% tumor content mix. Variants at and above 0.2% VAF are shown.

CONCLUSIONS

In support of science for a safer world, we have created a set of reference materials that can be used to assess the sensitivity and specificity of patient-specific MRD assays at defined VAFs.

When A>C and C>G somatic variants are used – e.g., those identified through WES – then the background appears to be close to 0.01% VAF with limited to no error correction (and avoiding C>A, C>T, and A>G may be helpful). This suggests that 0.2% tumor content sensitivity may be feasible without error correction and that 0.02% tumor content (or even lower) sensitivity may be attainable with additional error correction and sequencing using about 20 somatic variants per sample. On a 4-color HiSeq with patterned flow cells, A>T somatic variants also appear to have a relatively low background. We do not know if the same holds true for 2-color platforms.