

Detecting myeloid malignancy minimal residual disease

Recent findings and laboratory considerations for post-treatment monitoring

Charna Albert

October 2021—Detecting leukemic cells for post-treatment monitoring in normal karyotype acute myeloid leukemia is challenging, but new approaches to minimal residual disease monitoring may make it increasingly possible in the clinical laboratory, David Wu, MD, PhD, said in an AMP webinar he presented recently on myeloid malignancy minimal residual disease detection.

Although there are well-established approaches for monitoring AML with recurrent translocations, post-treatment monitoring of normal karyotype AML (aside from *NPM1*-mutated AML) has been more difficult because of the diversity of mutations present and has often focused on flow cytometry. Now, with broad implementation of next-generation sequencing platforms, there is interest in applying sequencing to detect MRD to improve standardization, sensitivity, and specificity.

It is an opportune time to consider this, says Dr. Wu, associate professor, Department of Laboratory Medicine and Pathology, University of Washington. In the past decade, numerous studies from investigators have annotated the substantial inter- and intra-patient clonal heterogeneity present in AML. In this regard, “we have a very detailed map of the genomic alterations that drive acute myeloid leukemia,” he says. Although adult and pediatric myeloid leukemias share many genomic alterations, there are age-specific differences in AML driver mutations. Recent findings reported by Hamid Bolouri, PhD, and Soheil Meshinchi, MD, PhD, and colleagues, of Fred Hutchinson Cancer Research Center, show that pediatric AML patients as a group are more likely to have chromosomal alterations driving their leukemia (Bolouri H, et al. *Nat Med*. 2018;24[1]:103-112).

For example, infants with AML often have chromosomal translocations involving the gene *KMT2A*. In childhood and young adulthood, core binding factor fusions such as *RUNX1-RUNX1T1* and *inv(16)* appear increasingly commonly. And in adult AML patients, there is generally a different proportion and profile of these genomic alterations. “Appreciating the diversity of these genomic alterations in AML is obviously critical when attempting to develop a comprehensive lab approach for post-treatment disease monitoring,” Dr. Wu says.

Strategic design of an MRD assay is a requirement not only for optimal testing but also for accurate interpretation and reporting of genomic alterations detected in the post-treatment context. For example, Dr. Wu says, not every gene variant correlates with an imminent potential for leukemia relapse. In adults, age-related clonal hematopoiesis (CHIP, for clonal hematopoiesis of indeterminate potential) can confound interpretation of molecular MRD, as CHIP mutations are common by age 50, even in younger, healthy adults without overt hematologic disorders (Xie M, et al. *Nat Med*. 2014;20[12]:1472-1478; Genovese G, et al. *N Engl J Med*. 2014;371[26]:2477-2487; Jaiswal S, et al. *N Engl J Med*. 2014;371[26]:2488-2498; Young AL, et al. *Nat Commun*. 2016;7:12484—the latter using error-corrected high-sensitivity NGS). “Clonal hematopoietic mutations in the blood are quite ubiquitous, particularly as we age,” Dr. Wu says. “The most common CHIP clones affect the genes *DNMT3A*, *TET2*, and *ASXL1*, but can also affect other important genes,” such as *TP53* and *JAK2*. These CHIP clones can and will be detected pre- and post-treatment, and even after the blasts clear from the patient’s blood or bone marrow, Dr. Wu says, and thus may confound data interpretation.

Similarly, because mutations in genes can be present in more differentiated cell types and not just leukemic blasts, interpreting NGS results in the post-treatment context can be tricky, he says. Recent single-cell RNA sequencing studies and mathematical modeling, in the lab of Brian Druker, MD, have shown that contrary to the dogma of AML being caused by a complete differentiation block, even a partial block in differentiation can lead to an accumulation of blasts (Agarwal A, et al. *Proc Natl Acad Sci USA*. 2019;116[49]:24593-24599). In this context, Dr. Wu says, because these differentiated cells still carry many of the same mutations present in blasts, this can confound data interpretation, as the VAF may no longer correlate with the proportion of residual leukemic blasts.

As their study showed, he says, persistence of genomic alterations, including both recurrent AML translocations and gene mutations, can be seen in all subtypes of AML. “This can impact pathologists’ interpretation of what these mutations may be in the post-treatment context.”

This issue similarly arises in the post-treatment context in which leukemic blasts are specifically induced to differentiate, Dr. Wu says. Akin to persistence of the *PML-RARA* fusion early post-treatment for acute promyelocytic leukemia, gene mutations identified at diagnosis in normal karyotype AML may persist even as blasts clear, as was observed in the *IDH2* inhibitor trial (Amatangelo MD, et al. *Blood*. 2017;130[6]:732-741). The variant allele frequency of mutant *IDH2* may thus not always track with the proportion of leukemic blasts. “In this context, NGS data interpretation requires careful correlation with the timing and treatment given,” he says.

As with any laboratory test, numerous issues determine the success and accuracy of MRD testing, Dr. Wu says. Sampling for myeloid MRD involves taking a very small proportion of bone marrow or blood from the patient. Given the limited sample, as the leukemia burden decreases in response to treatment, the potential for false-negative results increases. Further, residual leukemia cells, especially at low levels for MRD testing, are not evenly distributed within the bone marrow, he says, but often are found in discrete, “niche” regions, which can confound testing due to sampling considerations. Sampling at multiple time points to assess the kinetics of these variants could be considered. “For example, one could require repeat testing to confirm initial findings, but such additional testing can be costly,” Dr. Wu says, “and currently the optimal interval for sequential MRD testing using next-generation sequencing methods has not been established, though studies are increasingly examining this question.”

Dr. Wu highlighted the important issues to consider in bringing on any molecular MRD assay, among them ensuring minimal carryover contamination. In the clinical laboratory, when samples may come from patients at various time points during their care, “the challenge is to ensure carryover contamination is negligible between diagnostic and frank relapse samples where there is a high blast count and high mutation burden versus those samples for other patients who have very low level or no evidence of disease.” For NGS applications, Dr. Wu says, “this includes, for example, ensuring that sample index reagents, used to label and identify patient samples during sequencing, are of appropriate stringent quality so that DNA sequencing reads from one patient do not inadvertently get bioinformatically assigned to other patients, as batched sequencing is necessary to make this type of testing cost permissive.” Numerous groups have studied this issue and showed the importance of addressing these critical workflow issues (for example, Bartram J, et al. *J Mol Diagn*. 2016;18[4]:494-506).

Labs interested in performing myeloid MRD testing will likely require a significant investment in resources and effort. In contrast to routine approaches for diagnostic evaluation of solid tumors by NGS in which labs typically report mutations at variant allelic frequencies of approximately five percent with 250x coverage per AMP/ASCO/CAP guidelines, much deeper sequencing is typically required for MRD applications. “Along with this,” Dr. Wu says, “there is an increasing need to reliably discern true positive from false-positive variants.” For MRD testing, “ideally one hopes currently to achieve detection of variants down to 0.1 percent VAF and below. For such testing, this means,” he says, “the initial pull of a bone marrow sample is ideal to minimize hemodilution by peripheral blood that may occur on later pulls.” Further, one has to be able to allocate enough sequencing reads to query all of those cells’ DNA, he adds. Depending on the size of a targeted NGS panel and a lab’s clinical test volume, “it can be challenging to have the requisite sequencing capacity to perform these tests, not to mention the bioinformatic capabilities to interpret and then report these data.” In short, he says, MRD testing by NGS can be significantly more complex than the routine NGS testing many laboratories currently perform.

When using NGS for detecting myeloid MRD, the laboratory approach for testing is different, Dr. Wu says. Most standard NGS instruments have error rates in the one percent range, and “for this reason, many clinical labs operate above this range, reporting a limit of detection in the approximately three to five percent range.” To detect variants at much lower variant allelic frequencies, enhanced wet-lab or bioinformatic approaches need to be considered, he says. Two general approaches have been reported in the literature: One is exclusively bioinformatic in nature, and the other uses unique molecular indexes (UMIs) on the wet-lab side for library preparation to correct

error. Both allow identification of low-level VAFs down to 10^{-3} to 10^{-4} or even lower, such as with duplex sequencing. When deep sequencing is needed for MRD, there will inevitably be an increasing proportion of spurious variant calls. “This issue is typically not of any relevance or concern for diagnostic testing when an assay’s limit of detection is typically around a five percent variant allelic frequency,” Dr. Wu says. “However, if one is attempting to perform post-treatment MRD testing for myeloid neoplasms, distinguishing very low level spurious error versus true mutations becomes critical.”

Position- or site-specific error modeling attempts to correct sequencing and PCR error by measuring for any nucleotide position how frequently the alternate three nucleotides are detected. Variants in tested samples that occur at a frequency above those expected distributions are then considered true positives, Dr. Wu says, noting this is an approach that some labs have taken already to decrease the variant limit of detection to below one percent. Work published in 2018 showed the potential of site-specific error correction approaches for NGS to add complementary value for post-treatment disease monitoring as compared with flow cytometry (Jongen-Lavrencic M, et al. *N Engl J Med*. 2018;378 [13]:1189-1199).

Two groups recently investigated ways to enhance this error modeling approach that may mitigate challenges in having to sequence at increasing depths with UMIs to achieve higher sensitivity. These studies employed context-dependent error modeling in which error models were created based on considering the adjacent nucleotide sequence context of the targeted region (Abelson S, et al. *Sci Adv*. 2020;6[50]:eabe3722; Ma X, et al. *Genome Biol*. 2019;20[1]:50). “By analyzing the adjacent bases of a given variant position,” Dr. Wu says, “the authors were better able to discern low VAF true mutations.” As Abelson, et al., detail in their work, “compared with other error suppression techniques, their bioinformatic approach demonstrated lower numbers of false-positive mutation calls and greater sensitivity,” Dr. Wu says.

Another approach for error correction, in the context of MRD monitoring, uses unique molecular indexes as part of library preparation to identify and discard errors. “UMIs are short oligonucleotide sequences—often six to 20 base pairs—that are tagged into the patient’s DNA template. Upon amplification, these UMI tags are then used to bioinformatically collapse sequence information together into consensus reads, after removing sequencing or PCR error, defined as variants that occur in some but not all of the UMI-tagged reads,” Dr. Wu says. Conceptually, consensus sequencing can occur as single-strand or dual-strand approaches. In both approaches, UMI-barcoding is used to remove the spurious errors that occur during PCR amplification and sequencing. MRD analysis and quantitation subsequently proceeds from using only these consensus reads for defining tumor mutation allelic frequency.

Several approaches have been described for consensus sequencing. For single-strand approaches, the earliest was the SafeSeq approach (Kinde I, et al. *Proc Natl Acad Sci USA*. 2011;108[23]:9530-9535). Another similar approach used molecular inversion probes to capture targeted regions for amplification, sequencing, and removal of errors (Hiatt JB, et al. *Genome Res*. 2013;23[5]:843-854). Dr. Wu and his colleagues (Waalkes A, et al. *Haematologica*. 2017;102[9]:1549-1557) and others (Thol F, et al. *Blood*. 2018;132[16]:1703-1713; Hourigan C, et al. *J Clin Oncol*. 2020;38[12]:1273-1283; Patkar N, et al. *Leukemia*. 2021;35[5]:1392-1404) pursued application of these UMI-error correction methods to MRD detection.

By contrast, dual-strand or so-called duplex consensus sequencing, as pioneered by Lawrence Loeb, MD, PhD, and colleagues at the University of Washington, considers evaluation of both strands of a DNA molecule and requires the variant (and its complement) to be seen in both strands (Schmitt MW, et al. *Proc Natl Acad Sci USA*. 2012;109 [36]:14508-14513). In this regard, Dr. Wu says, “duplex sequencing is more stringent and thus more accurate than single-strand consensus sequencing in that it seeks to exclude early cycle PCR error by specifically tagging the paired Watson/Crick strands of the DNA template to confirm that a mutation and its complement are indeed both seen on the two strands.” If a variant is observed on only one strand of the DNA template, but not its complement, then this variant is interpreted to be an error and is discarded from further consideration. Due to this additional stringency for defining a true variant, duplex consensus sequencing is considerably more accurate for identifying extremely low-level VAF variants well below 0.1 percent (one in 1,000) and can reach one in 1 million, Dr. Wu says.

"The cost, however, is that one has to sequence more deeply to identify the paired sequences, which may be more challenging to achieve."

Next-generation sequencing is just one of the several methods being applied now for MRD detection. Conventional approaches such as RT-PCR are being complemented with newer approaches, such as RNA-Seq for fusion detection (Dillon LW, et al. *Haematologica*. 2019;104[2]:297-304), Droplet Digital PCR, and error-corrected methods of sequencing, as well as ultra-sensitive chimerism assessment as a surrogate for relapse and evaluation of circulating tumor DNA. While these methods have largely been tested in research studies, Dr. Wu says, "there is obvious interest in applying these technologies in the clinical realm moving forward."

Detection of circulating tumor DNA is emerging as a potential biomarker that has now also been explored for hematopoietic neoplasms, including AML. For example, Sousuke Nakamura, MD, and colleagues of the University of Tokyo recently demonstrated in AML/MDS patients who underwent stem cell transplantation that "in fact you can achieve the same sensitivity as bone marrow testing," Dr. Wu says (Nakamura S, et al. *Blood*. 2019; 133[25]:2682-2695). In their work, the authors collected tumor and available matched serum samples from 53 patients at diagnosis and post-transplant. After identifying driver mutations in 51 patients using NGS, the authors then designed at least one patient-specific Droplet Digital PCR assay for each patient. Diagnostic ctDNA and matched tumor DNA exhibited excellent correlations with variant allele frequencies upon testing, and both mutation persistence in bone marrow post-allogeneic stem cell transplantation and corresponding ctDNA persistence in the matched serum were comparably associated with higher three-year cumulative incidence of relapse. This approach thus appears promising, Dr. Wu says, and could be advantageous for patients, particularly for those who may have poor marrow cellularity and low blood count recovery post-treatment.

Another emerging approach is the use of blocker displacement amplification probes to enhance detection of variant alleles, as developed by David Zhang, PhD, and his research group at Rice University. Dr. Zhang's strategy is based on blocking amplification of the wild-type allele, resulting in potential variant enrichment by several hundredfold to enable rare variant detection below 0.1 percent VAF, using low read-depth sequencing on the order of about 300× coverage, versus the higher depth of coverage typically needed for consensus sequencing-based approaches. In this work, Dr. Zhang's group showed the detection of single nucleotide polymorphisms with a VAF of approximately 0.02 percent in a multiplexed panel with limited sequencing coverage (Song P, et al. *Nat Biomed Eng*. 2021;5[7]:690-701). Though this approach can readily enhance sequencing of multiplex hotspots using a low-depth sequencing method, Dr. Wu says, it may be more difficult to target the full coding regions of genes such as needed for some genes like *TP53*. "This approach nevertheless is likely one to have potential clinical relevance," Dr. Wu says, "as many clinical labs do not typically have the scale of DNA sequencers that can achieve the depth of coverage needed for sequencing a comprehensive panel of AML gene targets for MRD testing."

Evaluation using Droplet Digital PCR is another approach for MRD detection. ddPCR is a highly accurate molecular approach that uses microfluidics to partition a sample into tens of thousands of discrete reaction chambers for PCR analysis and subsequent discretized detection and quantitation. A notable advantage of ddPCR, Dr. Wu says, is its reliance on using Poisson counting statistics for quantitation of rare events, and thus it does not require external standards for quantitation as does RT-qPCR.

Overview of MRD methods

Established
<ul style="list-style-type: none"> ■ Reverse transcriptase quantitative PCR (RT-PCR) ■ Droplet Digital PCR (ddPCR) ■ Next-generation sequencing (NGS) <ul style="list-style-type: none"> Error modeling Error correction (unique molecular indices) Duplex sequencing ■ Chimerism ■ Multiparametric flow cytometry, FISH/cytogenetics
Emerging
<ul style="list-style-type: none"> ■ ctDNA ■ Enhanced bioinformatic error modeling ■ Blocker displacement amplification ■ Single-cell sequencing (combined genomics/proteomics)

Many in the field, including investigators at the University of Michigan, have shown the potential of using ddPCR to monitor patients post-therapy in AML with a limit of detection reported in their work as low as 0.002 percent VAF, he says (Parkin B, et al. *J Clin. Invest.* 2017;127[9]:3484–3495). The authors used patient-specific assays targeting on average about two to three mutations per patient and showed the potential to detect clones at very low levels. ddPCR is an important platform for labs to consider, Dr. Wu says, adding it's an approach that is somewhat constrained, however, "by the fact that only a few targets can be tested simultaneously, and as such it's harder to conceive of developing a broad, multigene panel-based test for MRD testing using ddPCR technology alone." Nevertheless, he says, clinical labs have developed assays using ddPCR to target frequent gene mutations, such as in *NPM1*, a gene mutated in nearly 20 to 25 percent of normal karyotype AML (Mencia-Trinchant N, et al. *J Mol Diagn.* 2017;19[4]:537–548), as well as in *IDH1* and *IDH2* (Ferret Y, et al. *Haematologica.* 2018;103[5]:822–829).

As another approach for MRD detection, labs have increasingly turned toward sensitive molecular tests to assess chimerism for patients who have undergone stem cell transplantation as a way to monitor engraftment (Khan F, et al. *Bone Marrow Transplant.* 2004;34[1]:1–12). These approaches are limited to patients who are post-allogeneic transplantation. "In this approach, molecular assays are designed to target and quantitate discordant alleles—either single nucleotide variants or other insertion/deletion mutations or copy number polymorphisms—that differ between the patient and donor. In this way, quantitation of the host cells or donor cells may inform the status of the stem cell engraftment and serve as a surrogate for leukemia relapse," Dr. Wu says, noting various groups have achieved this using RT-qPCR approaches. More recent work by some, including by Dr. Wu and UW colleague Stephen Salipante, MD, PhD, used single-molecule molecular inversion probes and NGS to target deletion copy number polymorphisms (Wu D, et al. *Clin Chem.* 2018;64[6]:938–949). Dr. Wu and others envision such ultrasensitive chimerism tests as complements to other approaches for MRD assessment.

Lastly, for myeloid MRD monitoring, NGS detection of insertion/deletion mutations can be sensitive without a need for significant bioinformatic or technical effort. Unique to insertion/deletion mutations, such as in *NPM1* and *FLT3*, is that the background error profile by NGS is quite clean, so that deep sequencing of these specific gene mutations can be performed without the need to use complex error correction approaches, such as is required for detecting SNVs. Currently, many labs use RT-PCR approaches for detecting *NPM1* gene mutation (a 4-base pair insertion), as highlighted in the seminal study in which an RT-qPCR approach was used (Ivey A, et al. *N Engl J Med.* 2016;374[5]:422–433). However, NGS approaches can detect this same *NPM1* insertion mutation, Dr. Wu says, "because the common *NPM1* 4-base pair insertion mutation does not commonly occur as sequencing or PCR artifact by chance." An advantage of an NGS approach for detecting *NPM1* mutation is that one can monitor MRD without a priori knowledge of *NPM1* allele and therefore can capture all of the different *NPM1* mutations, as well as assess clonal evolution (Thol F, et al. *Genes Chromosomes Cancer.* 2012;51[7]:689–695; Salipante SJ, et al. *Mod Pathol.* 2014;27[11]:1438–1446; Bacher U, et al. *Haematologica.* 2018;103[10]:e486–e488).

As several groups have also shown, the ability to deeply sequence *NPM1* mutation allows for potential MRD monitoring in a substantial proportion of normal karyotype AML patients, similar to RT-PCR (Patkar N, et al. *Oncotarget*. 2018;9[93]:36613–36624; Ritterhouse LL, et al. *Mol Diagn Ther*. 2019;23[6]:791–802). A comparable approach for detecting *FLT3*-internal tandem duplication mutations using highly sensitive NGS has been described (Blatte TJ, et al. *Leukemia*. 2019;33[10]:2535–2539).

As per the 2018 European LeukemiaNet Working Party guidelines, post-treatment testing for MRD is now standard of care in AML. Patients with mutant *NPM1*, *RUNX1-RUNX1T1*, *CBFB-MYH11*, or *PML-RARA* typically should have molecular testing for post-treatment monitoring. For other subtypes of AML, particularly normal karyotype AML, flow cytometry is commonly used. “It is hopeful that next-generation sequencing can play an increasing role,” Dr. Wu says. The field still has important work to do, with clinical colleagues, to ensure assay performance including accuracy of results, to define the clinical validity of reported variants, and to optimize time-points for testing. Many groups worldwide are advancing these efforts, Dr. Wu says, citing a recent review (Yuest JM, et al. *Front Cell Dev Biol*. 2020;8:249).

While one goal of clinical testing could be to detect only those mutations present at AML diagnosis, Dr. Wu’s view is that NGS is likely to be used “to detect any variant clone reliably in the post-treatment context using generic panel-based tests.” The challenge, he says, is developing an appropriate lab infrastructure to sequence broadly (as many AML genes as possible) and deeply enough (beyond a VAF of 0.1 percent), while minimizing false-positives and defining the clinical significance of variants that are most likely to correlate with imminent risk for disease relapse in a relevant clinical time frame.

“And all of this has to be done with a reasonable turnaround time,” he says, “and with the typical challenges of cost and oftentimes a lagging reimbursement landscape.”□

Charna Albert is CAP TODAY associate contributing editor.