

# Time now for tumor mutational burden?

## Karen Titus

November 2018—Like a piece of so-called sticky music, cutoff numbers can persist in physicians' minds outside of any real clinical value and, in the process, leave their laboratory colleagues mildly befuddled (not to mention searching for more useful cutoffs).

Such a jingle is creeping into tumor mutational burden. Lauren Ritterhouse, MD, PhD, co-director of the clinical genomics laboratory at the University of Chicago, recalls early conversations about TMB at her institution. Amid discussions about how and when to implement the testing, one colleague announced to all assembled that the cutoff number should be 100.



Dr. Lauren Ritterhouse at the University of Chicago, where the clinical genomics laboratory she co-directs has a research module for measuring tumor mutational burden. "Our data actually looks close to perfect," she says. [Photo by Bruce Powell]

"I asked, 'Do you mean 100 per exome?'" she recalls. The colleague was unsure—but kept repeating the number anyway: *I don't know. But I know it's 100.* As it turns out, says Dr. Ritterhouse, who is also assistant professor of pathology, the figure had some basis in fact, if not usefulness. One of the initial publications on TMB measurements used 100 mutations/exome as a cutoff value, and that's what stuck in her colleague's head.

More recently, another number has emerged, based on the use of the FoundationOne CDx assay (Foundation Medicine) in a recent clinical trial. "My colleagues are seeing the number 10—that's a number that's been thrown around lately. But it's just a number," Dr. Ritterhouse says.

Likewise, TMB is just a test—but one drawing more attention for its potential use as a predictive biomarker in immunotherapy, particularly in non-small cell lung cancer, melanoma, and urothelial cancers.

The idea is basic, says Alain Borczuk, MD: Tumors eventually learn to evade the immune system following initial attacks. But subsequent immune system response may be more robust in cases where the tumor expresses neoantigens on the cell surface, says Dr. Borczuk, vice chair of anatomic pathology and professor of pathology, Weill Cornell Medicine. When the immune system is reactivated by immuno-oncology drugs, the response will be

better if the tumor is more immunogenic; TMB measures the number of mutations within a tumor genome.

“It’s a theory,” says Dr. Borczuk. The cancer field is full of theories, of course. But this one has some support, he says, although it’s not uniformly recognized that TMB is the only measure of response to these drugs, also known as immune checkpoint inhibitors.

Different approaches have been used to try to predict response, including, most prominently, PD-L1, which has become the standard biomarker in a variety of tumor types. “Everyone recognizes it as an imperfect biomarker, but nevertheless it’s the one we have,” says Dr. Borczuk. Perhaps, goes the hope, TMB will be a better or potentially a different predictor for a set of patients who might respond to immunotherapy. This hope grew bigger on the basis of a study involving an examination of tumors that had microsatellite instability, or MSI, which seemed to indicate a better response to the immuno-oncology drug pembrolizumab, regardless of site of tumor origin.



TMB testing is ‘not going away. It is a piece of data that people will want.’

— **Alain Borczuk, MD**

With TMB thus nudged into the spotlight, compelling data have emerged, primarily related to NSCLC, specifically from the CheckMate 227 study, which looked at TMB and PD-L1 in patients who were offered either chemotherapy, an anti-PD-1 drug, or an anti-PD-1 drug in combination with an anti-CTLA-4 drug.

Reports Dr. Borczuk: “What they found—and this is the data that has been most discussed—is that patients who had high tumor mutation burden had a longer progression-free survival and a longer duration of that response with combination immunotherapy, when compared to chemotherapy.” It was, he says, “the type of difference that really captures your eye.” One-year progression-free survival with nivolumab plus ipilimumab was 42.6 percent, versus 13.2 percent with chemotherapy. Median progression-free survival was 7.2 months versus 5.5 months.

The study set the bar, he says, for TMB in predicting the subset of patients who could potentially respond to combination immunotherapy, especially since it seemed to be independent of the PD-L1 level. A second study, also looking at NSCLC, again found that higher TMB seemed to be a predictor of higher response rates to immuno-oncology.

Now comes the wait to see if the CheckMate 227 results will nudge the FDA to approve combination immunotherapy based on the biomarker.

In the meantime, laboratories can ponder the “hows” of TMB testing. Dr. Borczuk has told his colleagues that developing the testing is a must. “It’s not going away,” he says. Even if the importance of this particular piece of data isn’t fully clear, “it is a piece of data that people will want.”

At the University of Chicago, such efforts are already underway. Dr. Ritterhouse’s lab currently runs a large (1,213 genes) targeted panel for its cancer specimens, sizable enough, she says, that they can easily bring on larger mutational pattern metrics like TMB.

But validation issues are fraught. The original study and data suggesting the usefulness of TMB in immunotherapy were done using tumor and normal whole exome sequencing studies, “so that’s still kind of seen by many as the gold standard,” Dr. Ritterhouse says. All well and good, but validating to this particular gold standard “means that however many samples you’re going to run, you have to pay for two exomes in addition to whatever your panel is.”

Her lab had a set of cases that had already been run on its in-house panel; on about 30 of them, they were able to obtain or already had additional DNA as well as normal tissue, on which they performed whole exome sequencing. WES isn’t a routine step for them, however, which meant “we had to spend a lot of time getting the pipeline and variant calls for those whole exome sequencing analyses.” In fact, she says, that turned out to be the trickiest part of obtaining an accurate TMB. “It’s easy to add up variants, but making sure you have accurate variant calls” is a different matter.

UC does tumor-only sequencing; matched-normal sequencing, she says, is more expensive, and the logistics are tricky, including a separate consenting process, sample access issues, and potentially longer turnaround times. Adding to the complexity, UC has a separate, nonpathology lab that does inherited germline testing to determine cancer predisposition. Given these many moving parts, she and her colleagues wanted to see how accurate a tumor-only approach would be. If the results weren’t good enough, they’d then look to matched-normal sequencing.

For now, there’s no need to look further. Dr. Ritterhouse says she’s pleased with the results so far, calling them “really nice. Our data looks good.” The huge panel it uses—a little over 3 megabases—helps. “I know a lot of labs that have much smaller panels are facing a tougher time. It’s hard to get an accurate sampling of what you’d get in a whole exome sequencing if you’re only looking at 50 or so genes.”

Adding to her confidence, she notes that Foundation Medicine’s test also uses tumor-only sequencing. “So it’s not unreasonable to think” it’s a good approach, she says, before adding, “Although no one really knows how they filter their germline variants. They have a proprietary algorithm that’s used, and you can make guesses as to how they do it, but it’s a bit of a black box.”

It’s a point of consternation, to be sure. “We would all love to know—it would be helpful to many, many labs. But.” She pauses before uttering a pragmatic, Zen-like phrase that seems to burble up when pathologists talk about TMB: “This is the setting we’re working in.”

At UC, “We’re still struggling with how to best filter out germline SNPs,” Dr. Ritterhouse says. Like other labs, “We use population databases, but you know those aren’t perfect. You’re going to throw out some variants that are somatic, and then you’re going to miss a lot of private inherited SNPs that a patient might have.”

Currently the laboratory has a research module for measuring TMB, but the clinical launch may be another six to nine months away.

What have been the challenges so far? “One is coming up with the data you want to test,” says Dr. Ritterhouse. UC essentially did 60 whole exomes. But, she says, many labs might find that financially daunting. Some colleagues with whom she has spoken are choosing instead to validate against samples they already had in their own institution that had been tested at Foundation Medicine.

Not having an existing informatics pipeline for whole exome sequencing added to UC’s challenges, she continues; most cancer labs, in fact, don’t regularly do whole exome sequencing. There’s a lot of pseudogene noise, along with artifacts and signals. The laboratory will need to devote a large chunk of time (“months and months and months” is how she adds it up) making a WES bioinformatics pipeline, or have access to someone else’s.

Even then, it’s tricky. In filtering out the germline variants, she and her colleagues took a strict approach. “But it ends up not working so well on MSI cases and high-TMB cases. We found that a lot of the actual somatic variants are getting thrown out.” It was “almost perfect for everything below 20 mutations per megabase,” however.

The issue hasn't been fully resolved. "We're trying to come up with new ways to get rid of the germline variants that won't penalize some of these high-TMB cases," Dr. Ritterhouse says. Options include looking at inherited, common SNPs and variant allele frequency. "If they're adjacent to your variant, then you can make some Bayesian analyses as to whether this might be inherited or somatic."

Apart from this ongoing frustration of tossing out too many germline variants, she sounds pleased. "Everything else looks so beautiful," she says, bringing an artist's appreciation to the results. "Our data actually looks close to perfect."

At Brigham and Women's Hospital and the Dana-Farber Cancer Institute, Jonathan Nowak, MD, PhD, and colleagues have been running a targeted sequencing panel since 2013 that has covered, in different iterations, about 300 to, now, some 450 genes. Over the past year, they've used the data generated from this panel, along with looking at mutational patterns, to make calls on mismatch repair status across all tumor types, which became important when pembrolizumab was approved for advanced, MMR-deficient tumors in 2017.

In concert with deploying MMR analysis, "We also realized that another very worthwhile thing to measure would be tumor mutational burden," says Dr. Nowak, associate pathologist at Brigham and Women's and instructor of pathology, Harvard Medical School. "We know, of course, that MMR deficiency correlates reasonably well with an elevated TMB, although many tumors that have an elevated mutational burden aren't MMR deficient."

So in mid-2017, the lab also began reporting TMB for every tumor sequenced on its panel—"pretty much every solid tumor we see at the Dana-Farber, and some hematologic malignancies as well."

Early on, he says, they struggled with the best way to report. Providing only a number would have limited utility, he says, especially for uncommon tumors.

The solution? They built a reporting module that presents the TMB for the current tumor not only as the number of mutations per megabase but as a percentile, comparing it to previously sequenced tumors of the same type. "If we are sequencing our 500th colon cancer, for example, we'd be able to say that the mutational burden is 12, and that's in the top 86th percentile of all colon cancers we've sequenced so far." Additionally, tumors are also compared at a percentile level to all cases, regardless of tumor type, previously sequenced by the panel, which helps provide a context for TMB results in uncommon tumor types.

Figuring out how to report and classify results "is, honestly, almost the hardest part of this work," Dr. Nowak says. For institutions that plan to offer TMB, "that's an open question." Not everyone has the resources of the Dana-Farber, he acknowledges. "Our situation is unusual."

Yes, it is, says Dr. Ritterhouse. "For them, it's a fantastic way to do it. We can provide those numbers," but lacking a vast database of their own, "they won't have as much power." Many laboratories, she says, simply report a number and sidestep the larger issue. Interpretation could depend on tumor type or drug therapy and the combination being considered. UC hasn't yet decided how it will report.

As for the calculation, Dr. Nowak notes that almost every step of the bioinformatics pipeline as well as preanalytic variables can interact and cumulatively influence the TMB number. For laboratories that do whole exome sequencing, calculating TMB is "pretty trivial. Because you've sequenced all the genes, it takes some variables off the table," he says. But most labs will probably use a targeted panel, since this provides deeper coverage and faster turnaround times. But that raises another question: How concordant must those results be to WES? "Is it OK if you're within five or 10 percent of the TMB as estimated by whole exome sequencing?" he asks.

The other challenges are typical of any assay, including having an adequate sample. "What we have generally found," says Dr. Nowak, "is if there is an adequate amount of DNA available to perform our sequencing assay, it's not a challenge to calculate TMB, as long as the tumor content of the specimen also meets our threshold. For instance, our validation studies show that we need 50 nanograms of DNA from a specimen with at least 20 percent tumor content. We know from these validation studies that if we have a specimen that meets both of those criteria,

we can reproducibly generate the same sequencing results for that specimen.”

“You can always run into trouble if you have an inadequate or borderline specimen,” Dr. Nowak continues. “But I think the other big challenge for TMB is ensuring that there is sufficient tumor content for whatever sequencing depth your panel provides.” This varies from institution to institution and even within a single lab, depending on whether the lab performs amplicon-based or hybrid capture-based sequencing. In a specimen with too little tumor, despite an adequate amount of DNA, “you might end up with an artificially low TMB.”

The turnaround time for the assay at Dr. Nowak’s lab is typically two weeks and is limited by the NGS process itself rather than by any TMB-specific step(s). This includes a review for sample adequacy, DNA isolation, all the initial preparation steps for sequencing, the sequencing itself, pipeline analysis, and variant interpretation and report generation. “I don’t see any shortcuts that might tell you about TMB without actually sequencing the genes themselves,” he says.

Meanwhile, clinicians are filling out their own dance cards a little differently. At UC, Dr. Ritterhouse says her clinical colleagues devote very little time to discussing cutoffs. Or, as she puts it, “They don’t worry about it as much as I do.”



Figuring out how  
to  
report and classify  
results  
'is, honestly,  
almost  
the hardest part  
of  
this work.'

— **Jonathan  
Nowak, MD, PhD**

Some are more interested in TMB’s negative predictive value, Dr. Ritterhouse says. “I’ve gotten that from quite a few oncologists.” In a PD-L1-negative case, for example, a negative TMB might steer them away from using immunotherapy, particularly if the patient has comorbidities, is elderly, etc.

At Brigham and Women’s, TMB results are generated automatically. Dr. Nowak says his oncologist colleagues showed little excitement when he first spoke to them about offering TMB results. If they couldn’t act on the number, they didn’t want to see it. A year in, however, many more oncologists appreciate having the information, “and it’s something we routinely discuss in our tumor boards.” Even though it’s not quite actionable yet, “we’re clearly starting to at least informally distinguish between tumors that have a higher or a lower mutational burden on average than we might expect for that tumor type.”

The conversations sometimes take a funny turn, he continues. “We’ll go to tumor boards, and they’ll look at a report and say, ‘Oh, this tumor has a mutational burden of 12—it’s a little high.’ And then they’ll look at another tumor [of the same type] that has a mutational burden of 8, and they say, ‘Oh, it’s less than 10—it’s low.’”

How quickly the new becomes norm. The time to unwind these so-called standards is now, before they get caught

in the brain like the Kars4Kids jingle. Says Dr. Nowak: "I always caution them that, based on the specifics of our assay, there is not a substantial difference between a tumor with a mutational burden of 8 versus 12."

Dr. Nowak foresees a time when clinicians might want to apply TMB to situations apart from predicting response to therapy, including the occasional diagnosis. "We've seen this repeatedly now in carcinomas of unknown primary." If such a tumor has an elevated mutational burden, it's often possible to look at its pattern of mutations to discern an underlying mutagenic process. "So if the tumor has an elevated TMB and is also MMR deficient, you might think this is a tumor that could have come from the colon," he says, whereas a tumor with an elevated TMB and a tobacco smoke signature could instead be suggestive of metastatic lung cancer.

"Occasionally, we have sort of the reverse example," he continues, "perhaps a squamous cell carcinoma of the lung that everyone assumes to be a primary lung cancer. Rarely, we've seen that those tumors will harbor a high mutational burden with an ultraviolet light exposure signature. And so that's actually strong evidence suggesting that the tumor has metastasized from a sun-exposed cutaneous site and is not a primary lung cancer at all."

While this goes a bit beyond measuring TMB, it reinforces the idea that an elevated TMB might prompt pathologists to consider the underlying "why." This might be helpful in only a small number of cases right now, Dr. Nowak says, but "when it is helpful, it is extremely helpful."

NGS panels are capturing this granular information within TMB, whether it's CTGA changes, dinucleotide changes, or frame shifts. Perhaps this can help determine therapy and provide hints to etiology, pathogenesis, and environmental exposures. It might even be more helpful, ultimately, to look at overall mutational spectrum and pattern, says Dr. Ritterhouse, rather than look only for a list of mutations, variants, or fusions.

Dr. Ritterhouse says the discussions with her colleagues have been lively if not definitive. The thoracic oncology department regularly requests TMB testing for research purposes. The director "would love for us to be reporting TMB," she says. And in general, "I think oncologists want every bit of data they can have, regardless of whether there's great evidence for it."

At the same time, she knows some physicians remain skeptical of the data suggesting TMB's utility. (More broadly, when talking to clinical colleagues from other institutions, she's found they want to see more data, such as overall survival.) That skepticism may turn out to be useful, in fact. "It's giving us a little extra time to figure this out." She welcomes this bit of breathing room. No one is saying they'll go elsewhere for testing if her lab isn't reporting TMB within, say, the end of the year.

And where would they go, exactly? Enter the Romulus and Remus of laboratory testing: standardization and validation. These two pillars, always important, have yet to be settled in TMB. In fact, the two key TMB studies relied on different tests and different cutoffs.

Every step in the TMB dance—and there are many—becomes more intricate across institutions. Ideally studies will have been done on the same sequencing platform, analyzed by the same pipeline, with the same reference genome and transcript settings, and analyzed the same tumor content. But this is a purity of line generally reserved for the Rockettes. In laboratories, "This is almost never the case," Dr. Nowak says.

Some algorithms look at essentially the entire DNA sequence of every gene that's coded. Other groups say it's too expensive and too inefficient to analyze such vast territory. It's possible, they say, to look at a subset of genes, albeit still millions of bases, and come up with a number that reasonably approximates the number obtained by looking at all the coding sequences.

But that raises the question of "how much genomic real estate you need to sequence before you have a number that fairly matches what you would see from whole exome sequencing," says Dr. Nowak. "Is it enough to sequence 100 genes? 200? 500? 1,000?" Moreover, for most targeted sequencing, the gene panels are not chosen at random; generally, specific genes are selected because they're important in cancer. "So you're sequencing a biased subset of all the genes you would be analyzing by whole exome sequencing." Could that introduce a bias

into the TMB calculation? “I think that’s a real possibility.”

For those interested in even greater detail, says Dr. Borczuk, there’s another consideration. “If you have to produce a protein that’s mutated in order to get an immune response, then only the mutations that result in a protein sequence change should matter,” he says. Again, the algorithms have varied from one paper to the next. “Some have been more inclusive, some less inclusive.”

The challenge of converting acceptable research test methods to a uniform test plagued PD-L1 testing, Dr. Borczuk says. Labs have to work hard to figure out if they are re-creating what was done in various studies, with their different antibodies and different platforms. He fears TMB will be messy as well. In a sense, it’s like rescuing lost choreography. You *think* this is what Martha Graham meant, but one is never quite sure if this is how the steps should look.

How does this problem get solved? “I’m not sure of the best answer,” Dr. Borczuk replies.

Since the drugs and the tests are proprietary, he says, “until the full FDA approval occurs, sharing of that is difficult. It’s limited by whatever rules have been put in place between the participants [pharmaceutical and diagnostics companies] in the clinical trials.” By the time pathologists outside the clinical trials become involved in the process—“We become much more engaged once the approval has come through,” he says—“we’re potentially six months to a year behind. Sometimes more.”

Timing is crucial, Dr. Ritterhouse agrees. She notes that several groups, including Friends of Cancer Research, are trying to standardize TMB. But she remains worried. By the time useful guidance emerges, “it may be too long after everyone has needed to report it.”

Dr. Borczuk’s biggest concern is that “we don’t have the biologically relevant samples to test. What we end up doing is test validation within the laboratory, but we can never do a full validation in a true cohort of responders and nonresponders. That’s a huge limitation.”

Fortunately, he says, matters have improved since the PD-L1 process. The major pharmaceutical companies “are asking what we think about things now,” he says. “And that’s a great help, because when it was being filtered only through their oncology contacts in medical centers, they were getting a skewed perspective of the scope of the problem.”

Nonetheless, for oncologists now bringing these drugs into clinical practice, “their instinct is to go with a commercial laboratory that has the most upfront brand for the test.” That means FoundationOne, the test used for CheckMate 227, becomes the de facto standard, even “if that’s not intentional.”

Looming larger is the issue that clinicians simply “want a number to work off of, and they want to justify why it’s valid,” Dr. Borczuk says. For labs, he continues, it’s often difficult after the fact to be able to say that “the number that you produce in your own laboratory is equally valid as the one that was used in the study trial.” At this point, he says, if FDA approval were given for nivolumab and ipilimumab for stage IV lung cancer, in the near term, he’d feel compelled to use Foundation’s test. “The same thing happened with PD-L1,” he recalls. “The day that was made public, we had to choose a reference lab” from among the small number doing the FDA-approved test.

He’s worried that laboratories will continue to fall down this rabbit hole again and again. “This seems to be the way it’s just going to happen every single time.”

The other key study used Memorial Sloan Kettering’s IMPACT panel. But those researchers did not address the combination immunotherapy scenario used in CheckMate 227. That could lead to a situation where combination immunotherapy would use the Foundation cutoff of 10, but a monotherapy with a laboratory-developed test would require a different cutoff. Dr. Borczuk sees potential for matters “spiraling out of control,” with every indication potentially having its own test and its own cutoff. “So this is a problem,” he says. “There’s no question.”

The impact of using different platforms, different bioinformatics, and different variant calling is real. Dr. Ritterhouse

says she and her UC colleagues used an outside laboratory to run the whole exome sequencing. “They did the variant calling and gave us the TMB number, and we did it separately.” The results gave them pause. “It was amazing—the exact same data, the same samples, the same sequencer—how vastly different the numbers were. And it’s just that their whole exome variant calling pipeline was nascent, shall we say, and not heavily tested. It was, in fact, wildly different. Wildly inaccurate.”

It was a useful exercise in demonstrating how all the so-called simple steps, such as variant calling, can make a big impact, she continues. Unfortunately, the way matters stand now, “it’s like every lab is trying to reinvent the wheel themselves.”

UC is part of a larger consortium, called GOAL, consisting of 17 academic institutions, which is trying to develop a consensus gene list that could be shared for cross-site, cross-testing methodology concordance studies. Demonstrating concordance between labs would be a solid step toward standardization, she says.

Beyond laboratory standardization, it will likely be important to harmonize reporting strategies as well. Dr. Nowak is grappling with “Do we need to dichotomize into low and high? Or low, intermediate, and high? Or do we need a more granular, quantitative breakdown?”

Given all the testing complications, the biology of TMB seems almost like a quaint afterthought. Nevertheless, says Dr. Nowak, “It’s probably important to think about this.”

How “active,” so to speak, is TMB? Does it change over the tumor’s lifespan? Dr. Borczuk says he suspects it does. But, he notes, current discussions revolve around late-stage tumors. Early-stage tumors might indeed have different burdens, but that’s not currently clinically relevant.

On a related note, however, pathologists should be testing the most recent sample—the one that documents stage IV disease—rather than an earlier one, perhaps from a resection, Dr. Borczuk suggests. Possibly down the road, if it does change over time, it might be a future marker for tumors that are becoming more aggressive, he posits.

Some research suggests that the real question isn’t necessarily if the TMB itself changes, but rather if a change in biology—possibly a mutation in the antigen processing and presentation pathway—enables the tumor to acquire a mechanism to essentially hide its neoantigens from the immune system. “You might still have a cell with a high mutational burden and many predicted neoantigens,” explains Dr. Nowak, “but maybe they never actually make it to the surface of the cell to be recognized.” At least for patients undergoing immunotherapy, he says, there could be “a strong selective pressure to somehow block your neoantigens from being expressed and recognized” on the cell surface.

TMB may not be equally useful in all tumors. Lung cancer, melanoma, and cutaneous squamous cell carcinoma, for example, all tend to have very high mutational burdens. Many pediatric tumors, on the other hand, which may be driven by translocations, often have very low TMB. But Dr. Nowak has also seen wide variation—over at least one order of magnitude, but sometimes several—within a single tumor type.

He suggests that the broader group of tumors that are MMR deficient will likely incur additional benefit from some type of TMB measurement. The option to give pembrolizumab is currently a binary decision. Is the tumor MMR deficient or proficient? But a number of sequencing studies have shown that not all MMR deficiency is equal, so to speak, in terms of mutational burden. Some MMR-deficient tumors have a TMB that’s very slightly above MMR-proficient or microsatellite-stable tumors. Others have 10 times as many mutations, or more. It seems reasonable, Dr. Nowak says, to expect that those tumors might respond differentially to immunotherapy and to have different prognoses. Studies looking at this are underway, which might show potential use of TMB as a marker to stratify within MMR-deficient or even MMR-proficient tumors.

In the meantime, TMB is, if nothing else, quietly making a run at becoming the Miss Congeniality of biomarkers. For any lab doing some type of genomic sequencing over sufficient genomic real estate, TMB values can be calculated fairly easily across all tumor types, says Dr. Nowak. “And almost every tumor has some number—it’s not going to



be zero.” That simplicity is appealing. “We don’t have too many markers that we can calculate and compare at a pan-tumor level.”□

*Karen Titus is CAP TODAY contributing editor and co-managing editor.*