AI-driven spatial biology: the next next-gen sequencing

Sherrie Rice

November 2023—Spatial biology may be an emerging field, but Kenneth Bloom, MD, says he and other pathologists have been doing it “since we got the microscope.” And he argues it’s going to become “the new, most important lens we look through.”

The reason is the emergence of new cancer treatments like immunotherapy and, most importantly, antibody drug conjugates like Enhertu, says Dr. Bloom, head of pathology for Nucleai, a company specializing in AI-powered spatial biology.

“Prior to these therapies, spatial biology was interesting—just one more way to classify disease,” Dr. Bloom said, speaking last spring at the Pathology Informatics Summit of the Association for Pathology Informatics. “The fact that we now have these drugs, and need to predict who’s going to get the maximal benefit from them, changes the game and requires us to develop new diagnostics that are beyond looking at tumor cells.”

Next-generation sequencing revolutionized the detection of multiple targets in one assay, facilitating targeted therapy, he says. NGS was about the need to “look at more and more targets in tumor cells” and having limited amounts of tissue. “How do we do that? Next-generation sequencing was a home run for that.”

Now, with immunotherapy and antibody drug conjugates, he says, it will be spatial biology that “takes us to that next-level biomarker” that will identify patients who will get the most benefit.

Spatial biology need not be restricted to spatial transcriptomics or multiplex immunofluorescence, he says. “Those are tools that are going to help us better understand things, and someday might be necessary in clinical practice. But the promise of machine learning and AI is taking those learnings and applying them back to materials we already have, like H&E and immunochemistry. And maybe we will need more complex assays over time.”

The goal is to get to 80 or 90 percent response rates in patients who have cancer, he says.

He calls targeted therapy a revolution, one that “sparked all of the things we’ve seen for the last 25 years.” But it’s “just one step,” he says, “and the reason it’s only one step is because while targeted therapy improved objective response rate and progression-free survival, it didn’t improve overall survival.” It improved quality of life, he says: better and faster response with fewer side effects.

“Good step forward, but it wasn’t the step required to actually cure cancer.”

How does spatial biology make 80 or 90 percent response rates a reality?

Pathologists have identified spatial characteristics for a long time, he says, citing as examples the infiltration of lymphocytes and how they migrate to and infiltrate tumors. “There are volumes and volumes written on observations we’ve made over time that haven’t been relevant until now.”

The simplest analogy is mapping, he says. A person with a paper map who needed to go from point A to point B could map the route. “But we all know now, with Apple Maps or Google Maps, you can ask all sorts of pertinent questions about the route. What’s the fastest way? The traffic patterns? What if I need to stop for gas? What if I need a charger? Now I can ask questions of that map and decide how I move.”

Fig. 1. Spatial analysis and NGS similarities

Step	NGS	Spatial platform
Sample	Multiple unstained sections	One or many stained sections
Library prep	Nucleic acid extracted, capture probes select areas of genome for analysis and extent of coverage.	Section is stained by H&E, IHC, or MIF. Number of slides, stains, and cells are analogous to genome coverage.
Sequencing	Platform chosen for sequencing and depth of sequencing determines sensitivity.	Platform is chosen for whole slide imaging. AI is used to determine cell types. The ability to do this accurately is similar to depth of sequencing.
VCF creation	Sequence is compared to a consensus normal sequence and variations are recorded.	Thousands of cellular and spatial features are extracted and recorded.
Annotation	Variant scientists review each variant recorded in the VCF and categorize them as benign, pathogenic, or VUS based on established knowledge bases.	Variant scientists use proprietary knowledge bases to identify features associated with, not associated with, or unsure association with outcome or hypothesis.
Reporting	Stand alone or integrate into existing LIS	Stand alone or integrate into existing LIS

Spatial analysis is beyond just mapping; it applies analytic methods to spatial data. In pathology, spatial analysis is multifold but begins with understanding every cell type in a slide and the relationships of those cells, Dr. Bloom says. “That’s problematic in itself because when you show a slide to a pathologist, we don’t look at it one cell at a time and say, ‘You’re a lymphocyte. You’re a dendritic cell. You’re a macrophage, and you’re a tumor cell.’ We don’t always agree on the diagnosis, let alone on individual cells that made up the diagnosis.”

AstraZeneca, maker of Enhertu, noted that its drug binds to cells that express HER2, but exerts its effect not only on that HER2-positive cell but also on adjacent cells, the so-called bystander effect. So AstraZeneca developed what it calls a spatial proximity score, which is the percentage of tumor cells that are HER2-positive themselves as well as those tumor cells that have at least one HER2-positive neighbor (within 50 μ).

Dr. Bloom

“Imagine trying to do that with just a microscope,” Dr. Bloom says. “You’re looking for positive tumor cells, and now you have to look for negative cells that are adjacent to a positive cell, but within a certain distance. We can barely evaluate low levels of HER2 expression and now you want us to also identify adjacent negatives?”

But that is what makes the difference, he says, in who responds to the drug and who does not—better than any other marker. “So spatial relationships make a difference with antibody drug conjugates.”

Spatial characteristics can be structured in a couple of ways, Dr. Bloom says.

One would be into regions. “Typical ways we as pathologists think of this would be, *Here’s the tumor, here’s the*

invasive margin, here's the stroma. We can subdivide the stroma as stroma next to tumor, stroma farther away from tumor. We also identify regions like vascularity, benign tissue, and necrosis. Whatever you come up with that you think is important to you can be an area structure."

Then the pathologist needs to identify and classify every cell within those regions. "If you rely on pathologist annotation to do that, you're stuck, because pathologists don't always agree on those annotations."

Pathologists are good at identifying some cell types like lymphocytes and tumor cells. "We're not so good at accurately identifying stromal cells, dendritic cells, and myeloid cells," he says. The spatial biology literature today is all about lymphocytes and tumor cells, "because we can identify those cells effectively."

It's trickier to develop machine-learning models that identify individual cells with the granularity pathologists need, "but then you can apply mathematical models to extract all the features and interrelationships of those cells. And tens of thousands of those can be extracted from a slide."

He argues that spatial biology has many similarities to next-generation sequencing (Fig. 1).

For NGS, what areas of the genome to cover has to be determined. Whole genome sequencing? Whole exome sequencing? Single genes like *EGFR*? Hotspots only? Exons? How far to go in the introns? "Those decisions influence the design of the assay, allowing you to answer your questions."

That holds for spatial biology too, he says. "Do you want to understand areas of the slide—stroma, tumor, benign areas, necrosis?" Do you want to understand cells—that it's a tumor cell, lymphocyte, macrophage, endothelial cell, fibroblast, and so on? "Or do you also want to know the cell state?" Whether it's an M1 or M2 macrophage? Whether it's a cytotoxic T cell, a helper cell, or a regulatory cell?

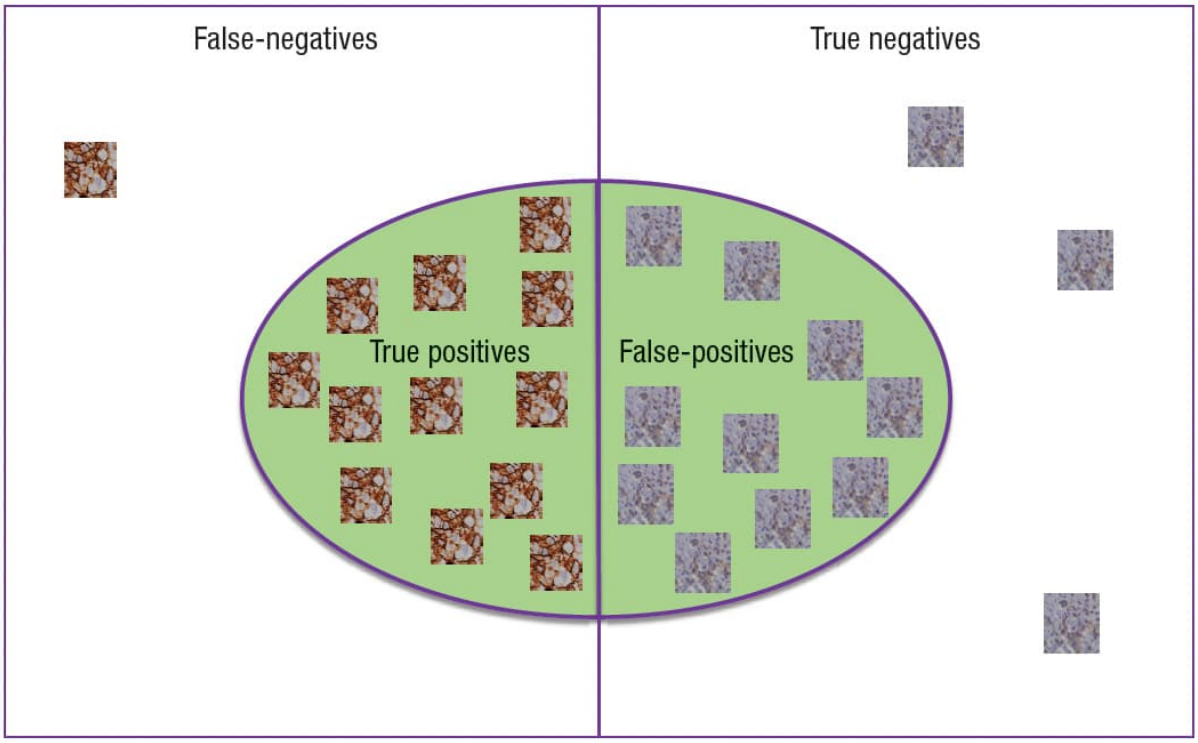
"All of that determines how you need to train your model. And I would argue it's the same as NGS," where the user has to know which panel is being run and what can be determined with that panel.

There's also a need to understand reliability. For NGS, depth of the read is important. Bidirectional or unidirectional? "When we find something, we want to understand what we found—was that chance or was that something that was actually there?" he says.

Spatial biology analysis frequently utilizes deep-learning algorithms, and truth can be determined in a number of ways. If pathologists annotate standard pathology slides, Dr. Bloom says, the models will be somewhat limited. "To really understand cells and cell states, you have to use multiplex immunofluorescence with a fairly large panel, single cell transcriptomics, or some other multiomic analysis."

How will a pathologist know if an AI model is useful? The metrics used to assess AI models are similar to positive and negative predictive value. The precision of a model is a measure of how many of the items retrieved by the model are relevant, and the recall is a measure of how many of the relevant items are retrieved. "Let's say we develop a model to predict HER2 positivity and we test the model on 1,000 tumors with known HER2 expression. The precision of our model measures how many of the 1,000 tumors predicted to be positive by the model are actually positive," Dr. Bloom says, "and the recall measures how many of the HER2-positive tumors are identified by the model."

In **Fig. 2** is an example of good recall (almost all of the relevant items are retrieved) but bad precision (few of the retrieved items were relevant).



In **Fig. 3** is the opposite: Of the retrieved items, “we got all the HER2s.” But of the relevant items, few were retrieved.

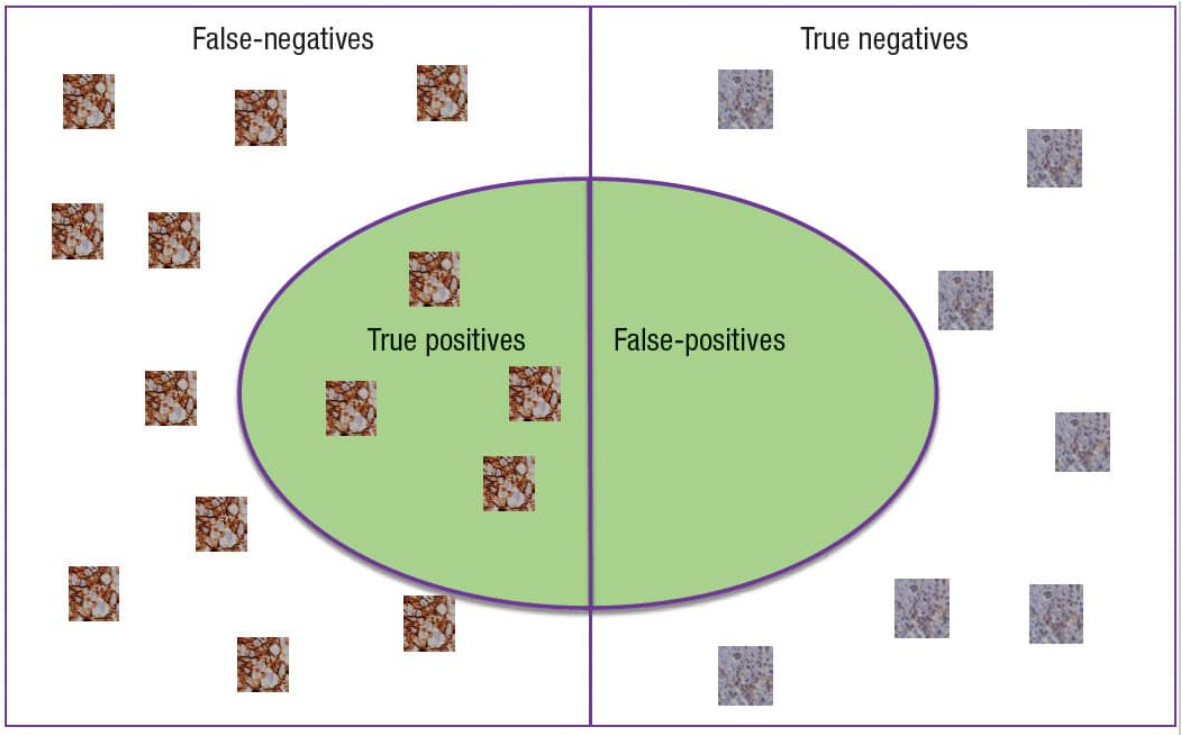


Fig. 3. Good precision but bad recall

- How many retrieved items are relevant (precision)
- How many relevant items are retrieved (recall)

An F1 score is used to measure the accuracy of a model. It’s also known as the harmonic mean: the reciprocal of the arithmetic mean, which can be calculated mathematically. It’s used as a measure to say the right elements have been captured and all the elements that should have been captured were captured. “The harmonic mean is the best measure of a model’s overall performance,” Dr. Bloom explains.

Many published deep-learning models have only moderate F1 scores, he says, ranging from 0.6–0.7, but these F1 scores are rarely reported, making it difficult for the reader to determine the value of the model. The higher the F1 score the more accurate the model.

Lastly, Dr. Bloom talked about variant calling.

For NGS, pipelines evolved over time. Initially, not all pipelines identified the same variants, and only certain variants, like single nucleotide alterations and small indels, could be reliably identified. Pipelines have now evolved to identify larger indels, copy number alterations, and structural variants. “For spatial biology, we are just at the beginning,” he says, “but we’re going to build similar robust pipelines” for biologic interpretation, pathology correlation, and cell associations.

Knowledge bases are essential to explain the clinical significance of identified variants, and it took many years before the first knowledge bases became available for NGS data. “For spatial biology, knowledge bases are in their infancy,” he says, “but they will develop over time.”

This is what’s needed, he summed up: Take the complex biological data that’s available, apply computational and machine-learning tools, develop the knowledge bases to create the insights, which will ultimately lead to the new model, which is health intelligence.

“It’s radically different from where we’ve been,” he says, “but this is where we’re moving. And I think you’re going to see this evolve over the next 10 years.”

Sherrie Rice is editor of CAP TODAY.