

“We both try to integrate our lab work and clinical work as much as possible.”



ASH ALIZADEH AND MAXIMILIAN DIEHN

# DNA DETECTIVES

There's no single formula for the elements of a productive partnership. But, as Ash Alizadeh and Maximilian Diehn would attest, a warm friendship certainly improves the chemistry. The Ludwig Stanford researchers have been fast friends for some 20 years now. "We're basically family," says Alizadeh. "We started medical school in Stanford at the same time. We were in graduate school together. We still work on the same lab bench, sharing reagents, food and all the stuff we did at graduate school. Our offices are right next to each other and we run joint group meetings. We're connected in almost everything we do."

That includes some rather brilliant science. In 2015, their extended collaboration resulted in the publication of two particularly noteworthy papers. First, their laboratories described in *Nature Methods* an analytical process named Cibersort that applies gene expression profiling and advanced computational analysis to trace back the precise spectrum of cell types contained in a slurry of disassembled tissue.

In the second study, published in *Nature Medicine*, Alizadeh, Diehn and their colleagues in the Center for Cancer Systems Biology at Stanford integrated the gene expression patterns of 39 types of cancer from nearly 18,000 cases with information on how long each patient survived. Their analysis of the resulting database, PRECOG, identified small

sets of genes that are associated—across a surprisingly broad spectrum of cancer types—with either good or bad patient prognoses. They then applied Cibersort to the cases they'd analyzed, discerning complex associations between patient survival and the presence of some 22 distinct types of immune cells in tumors.

## An eye on patients

Aside from pipettes, food and reagents, Alizadeh and Diehn also share a guiding principle. "We both try to integrate our lab work and clinical work as much as possible," says Diehn. Alizadeh, a clinical oncologist, specializes in lymphoma, while Diehn, a radiation oncologist, focuses on lung cancer. These are, respectively, among the most common blood cancer and solid-tumor types.

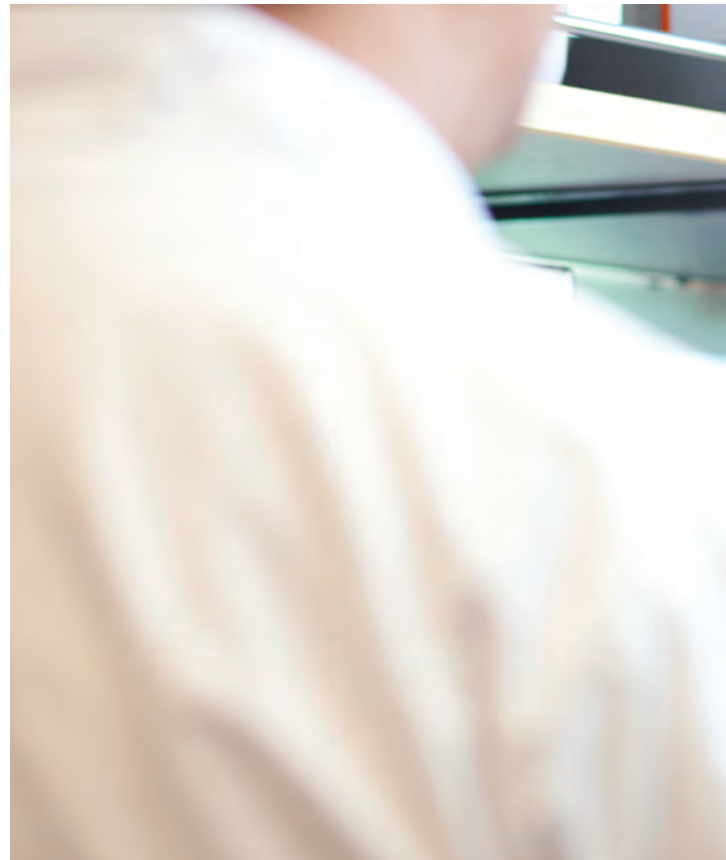
“We think what we’re working on in these areas should extend to cancer broadly, in line with Ludwig’s mission,” says Alizadeh. “So we try to take moonshot types of experiments and reduce them to practice, to what can be done for the average patient.”

Consider their work with circulating tumor DNA (ctDNA), which is shed by dead cancer cells and can, with some analytical finesse and the latest technologies, be detected in blood and other fluids. “We both became interested in ctDNA as part of our clinical routine,” says Diehn. “Ash was running some assays available for lymphoma patients, and I was frustrated that there are no good biochemical markers for lung cancer.”

In 2014, the pair reported in *Nature Medicine* a highly sensitive, minimally invasive method for detecting non-small cell lung cancers (NSCLCs) in patients. Their technique, CAPP-Seq, detected 50% of Stage I NSCLC, and 100% of Stage II-IV NSCLCs. The researchers also found that they could assess treatment responses earlier than they did using radiologic imaging. They are now developing more sensitive tests for Stage I NSCLC—when the cancer is most treatable—and similar tests for other malignancies, including those of the blood, brain and gastrointestinal tract.

These applications were enabled by improvements the researchers made to CAPP-Seq. Their new method, named integrated digital error suppression (iDES), combines two clever and complementary strategies to eliminate errors introduced when ctDNA is captured and prepared for sequencing and was reported in a March 2016 paper in *Nature Biotechnology*.

“Could we use it for screening? That, of course, is the Holy Grail and the hardest problem,” says Diehn. “But we could also use it to monitor response to drugs or the



development of drug resistance in patients. Those are just some of the projects we’re working on.”

### Hitting rewind

Cybersort likewise had its roots in a clinical problem.

The severity of a cancer is often tied to the diversity of cells in a tumor, and information on that diversity can be key to effective treatment. Pathologists and researchers get a handle on that diversity today through flow cytometry, in which cells in a sample are separated and then labeled and counted using antibodies, or via microscopy. This is relatively easy when dealing with blood. But tearing solid tumors apart can destroy certain types of cells and, in either case, the antibodies required for a comprehensive profile are not always available.



Ash Alizadeh  
Ludwig Stanford

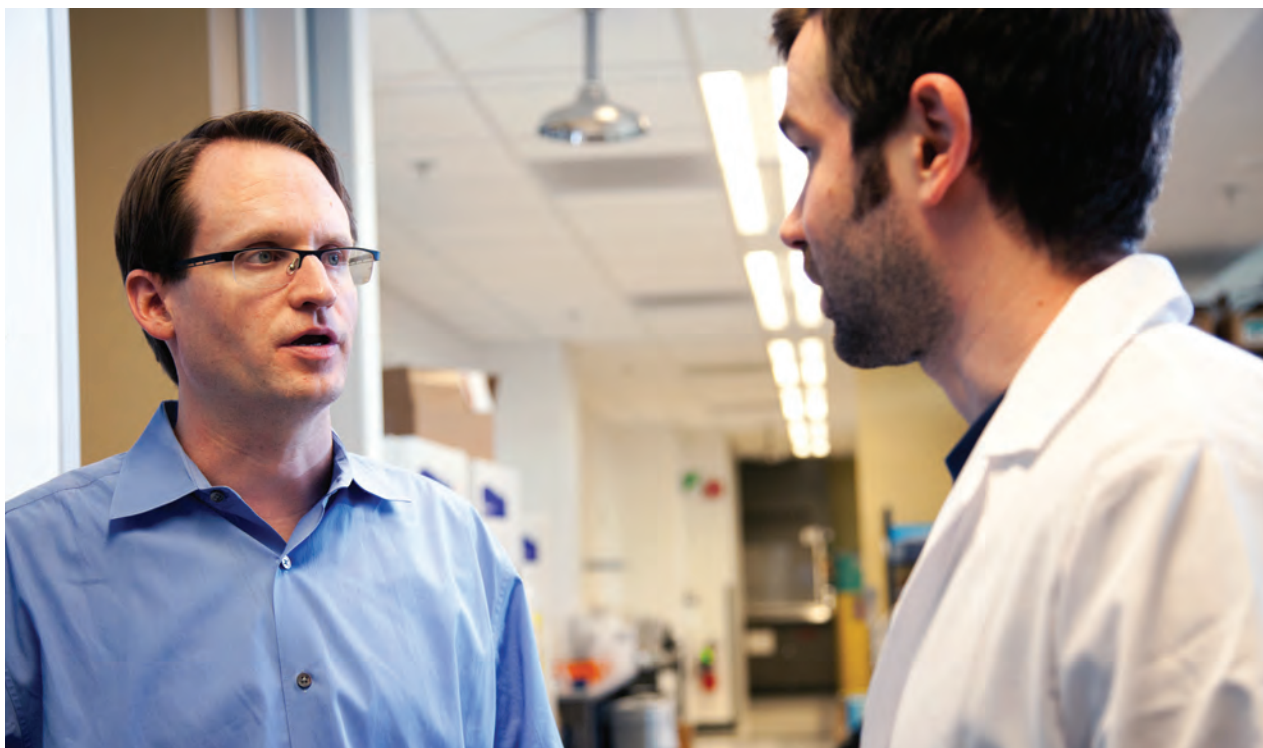
“When Aaron Newman, a very talented postdoc in our group, first approached me about trying to tackle this problem computationally, I was skeptical,” recalls Alizadeh. Newman was proposing to turn tissue into the equivalent of a smoothie and use the molecular clues in there, plus some software, to identify the cell types in the intact tissue. Others had tried similar approaches before with uneven results. “But he took up the challenge with such fierce and unwavering commitment that he produced something pretty powerful,” says Alizadeh.

One big challenge in any such analysis is that all cells express a basic subset of housekeeping genes, making for a lot of data “noise” through which the unique signal of a specific type of cell must be detected. Another is that any tissue sample is likely to

contain cell types the computer does not know about and so cannot take into account in its analysis.

To build Cibersort, the researchers read the transcripts of expressed genes from about 20 cell types and developed a fingerprint for each based on about 500 or so genes each characteristically expresses. Newman then applied machine-learning algorithms—the sort of programs used in speech recognition and self-driving cars—to address the anticipated data noise and confusion, and to reconstruct the tissue based on those fingerprints.

Cibersort requires much less work and introduces fewer variables than current methods for analyzing cell types in tumors, says Diehn. It is also sublimely precise. “We’ve been able to detect very closely



**Maximilian Diehn**  
Ludwig Stanford

related yet distinct subsets of immune cells, down to really low fractions of one percent or less of the total,” he notes. “We can even discern active subsets of immune cells from inactive ones.”

### **Good prognoses and bad**

PRECOG too addresses a sticky problem in cancer research. Researchers have not been able to figure out precisely how the profiles of genes expressed in tumors correspond to outcomes in most cancers. Many have found patterns, but such findings have largely been hard to replicate, except in a few types of malignancies like breast cancer. This, says Alizadeh, is because most such studies were too small compared with the number of genes expressed by cancer cells.

But about five years ago, the researchers noticed that there was a critical mass of relevant data available: Tumor samples from tens of thousands of patients had been

genetically profiled and stored along with their clinical outcomes.

Working with their colleagues at the Stanford Center for Cancer Systems Biology, including Andrew Gentles and Sylvia Plevritis, the researchers built a database that they named PRECOG by curating data from their own studies and those of their past collaborators, and by browsing information deposited in public repositories. The results of their analysis were surprising.

“Across the 39 cancer types we looked at, tumors had far more in common than they did distinguishing features when it came to patient prognosis,” says Alizadeh. “About two-thirds of the genes that are prognostic for one cancer are prognostic for at least one other cancer type.”

The team was able to identify the top 10 genes broadly associated with each prognosis. In particular, high expression

of *FOXM1*, a gene involved in cell growth, was associated with a poor prognosis across cancers. Meanwhile, the expression of a gene known to be involved in immune responses, *KLRB1*, seemed to have as broad a protective effect.


To get a better sense of the immune component of outcomes, the researchers applied Cibersort to the problem, painting a sweeping portrait of the association of immune cells with prognoses.

“Immune cell contributions can be upside-down from cancer to cancer,” says Alizadeh. “An immune cell like a macrophage can be favorably prognostic in a lymphoma, but very adversely so in, say, a breast cancer. If you were trying to engage macrophages with an immunotherapy, you might want to know about that.”

PRECOG, which is freely accessible, has obvious implications for cancer research and the development of new therapies and diagnostics. Diehn has already led a study, published in the *Journal of the National Cancer Institute* in 2015, reporting a potential diagnostic test for NSCLC based in part on information gleaned from PRECOG.

“The test predicts which patients will benefit from more aggressive, systemic therapy after having lung tumors removed in early stage lung cancers,” says Diehn. The test would have to be validated in a large clinical trial, he notes, but because it only requires detection of nine genes, most clinical labs would be able to perform the assessment.

“As oncologists,” says Alizadeh, “we are often humbled by the fact that we’re shooting in the dark and lack the tools we need to see the responses we’re hoping to see. But we’re very hopeful.”

With good reason, it would appear. 

**The researchers built the PRECOG database by curating data from their own studies and past collaborations, and with information in public repositories. The results were surprising.**