“Opening up Access to Cancer Data for Patients" (Frank Nothaft) [#76]

Brad Power
November 8, 2023

“My goal is to make it easier for patients to work with their diagnostic data and their EHR data and make empowered decisions about their care.” – Frank Nothaft

“You should get to a system that makes the data a lot more usable, that hopefully exposes data that is clean and that has clinically good interpretations, and the right cautions on how to interpret it.” – Frank Nothaft

Meeting Summary

Advanced cancer patients want access to as much data as possible to help them personalize their testing and treatment decisions. But cancer genomics data can be very hard to understand or access except for a small number of bioinformatics experts. The coming personalization revolution in cancer care and healthcare generally will depend on breaking down data barriers for patients so that they can uncover actionable insights and make empowered decisions about their care with their medical team. Diagnostic companies and data and AI services companies, especially startups, are potential partners who can accelerate access and personalization.

Frank Nothaft is uniquely qualified to explore the promise and challenges of data access for patients and caregivers. Following his PhD in computer science from UC Berkeley, he worked in bioinformatics at Databricks, a company focused on big data processing, where he was very focused on open source bioinformatics, especially around cancer genomics. He also worked at Tempus, the diagnostics company, where he worked with a much broader crowd of folks who had deep expertise in oncology, but less knowledge of bioinformatics. He realized how inaccessible cancer genomics data was to non-bioinformatics experts.

Frank left Tempus in mid-August 2023 and is starting to work toward solving this problem entrepreneurially. He is prototyping a solution that both makes it easier for a broad swath of people (patients, caregivers, clinicians, and researchers) to visualize and analyze genomic and other data, and that helps to expose information, e.g., about “the cancer genome”, in a digestible way -- personalized to the patient, telling them what they need to know in a way that is tailored to their listening, e.g., at an 8th grade level.

What is the promise of cancer patients and caregivers having easier access to information that will help them in their decision-making?

- **Personalization**: Blend what patients and caregivers are looking to get out of their care with what clinicians have seen in the clinic and evidence from research literature to pick a treatment that is most likely to meet the patient's goals for their treatment, including quality of life, in an effective and safe manner
- **Better ideas**: Uncover actionable insights to bring to their medical team

“Opening up Access to Cancer Data for Patients" (Frank Nothaft) [#76]

- **Empowered**: Make sure that patients can easily access the information they need to participate as co-pilots with their medical team in decisions

What are the barriers that patients and caregivers face in accessing and using information that could help them make better care decisions?

1. **Incorrect tests**: Some patients don’t get the diagnostic tests that might guide their treatment decisions (a data source for personalization).
2. **Test misinterpretation**: Many patients and physicians are unfamiliar with using novel tests (e.g., whole genome, RNA sequencing, AI-powered multimodal tests) to drive personalized treatment decisions.
3. **Hard to access, quickly changing medical knowledge**: Medical literature is vast and jargon-laden, making it difficult for patients and caregivers to use. There is so much available medical data, but as new tests and treatments are constantly emerging, making it hard to keep up with the latest research and developments and how patients actually fare in the real world on new treatments.
4. **Patient vs. system values misalignment**: Cancer research optimizes for overall survival and progression free survival, at the possible expense of quality of life.
5. **Insufficient information on side effects**: Patients need more information on the side effects of cancer treatments and their impact on quality-of-life. Having access to real-world incidence rates, as well as advice on how they can mitigate those side effects will help them get the best possible response to the medicine.
6. **Personalized vs. population evidence**: Information for personalized treatment decisions must take into account individual patient characteristics and preferences, which depend on real-world evidence from individual patients’ experiences, while the gold standard of evidence from randomized clinical trials provides population-level evidence.
7. **Physician persuasion**: When patients get information, especially from novel tests that guide them to treatment options outside the standard guidelines, they need to effectively share that information with their physicians to persuade them to prescribe the treatment they believe is best for them.
8. **Reconciling test results**: Reconciling conflicting data from multiple cancer diagnostic tests, e.g., between RNA sequencing and IHC for PDL1 protein in cancer diagnosis.
9. **Inconsistent diagnoses**: Different interpretations and recommendations from different physicians.

What are the possible solutions to these barriers?

1. **Up-to-date medical databases**: Patients and caregivers will access continuously updated medical knowledge bases (like WebMD, but curated to meet the specific needs of the rapidly evolving field of oncology) to make it easy to answer questions (including quality of life side effects), and provide the sources and experts to bring to their clinicians.
2. **-Omic databases**: Patients and caregivers will access large genomic databases (e.g., TCGA) to provide contextualization of -omic findings (e.g., gene upregulation), including normalizing for various biases.

3. **Observational registries with patient-reported outcomes**: Patients and caregivers will access observational registries to see the experiences of “patients like me” who are taking their treatment or tests or treatments they are considering.
4. **Data cleaning**: Software services will clean and normalize data (e.g., inclusion/exclusion criteria to improve clinical trial matching).
5. **Information filter**: Software services will continuously read all the medical research literature and provide patients with updates, including alerts, on new medical evidence, clinical guidelines, diagnostic tests, treatments, and trials that are important for them, and summarize it at a level they can understand.
6. **Interactive chatbot**: A personalized software agent (like ChatGPT) will use the curated medical knowledge base to enable patients and caregivers to ask questions about their diagnostic reports, treatment choices, or any other questions at a level they can understand.
7. **Communication channels:** Very complete, very well secured connections between institutions and with the patient and caregivers will enble the oncologist, pathologist, and the diagnostic testing company to share data to get to a better diagnostic readout.
8. **Models**: Longitudinal data analysis, algorithms, and AI will help understand the evolution and progression of cancer over time, the effects of multiple rounds of therapy, and predict treatment response, using evolutionary and game theory.
9. **System navigator role**: A system navigator will check information quality and help patients and caregivers interpret information and communicate with their medical team and advocate for their needs, especially where multiple doctors and diagnostic specialists are involved in a patient's care. They will provide timely and personalized feedback to patients (not as busy as physicians), avoiding confusion and frustration.

How can you engage if you are interested in discussing this topic?

You can join the online discussion with Frank at community.cancerpatientlab.org to answer questions such as:

1. What are your most important care decisions, and what information do you need to make those decisions?
2. What barriers make it hard for you to access and analyze your data?
3. How would you like to interact with and visualize your data?
4. Having arrived at ideas on what is best for you, what evidence do you need to bring to your medical team?

“Opening up Access to Cancer Data for Patients" (Frank Nothaft) [#76]

Meeting Notes

SUMMARY KEYWORDS

patients, data, diagnostic, frank, information, people, working, cancer, questions, company, understand, oncology, point, areas, diagnostic testing, side effects, rna, interpret, physicians, research

SPEAKERS

Frank Nothaft (58%), Brad Power (13%), Jeff Krolick (12%), Rick Stanton (8%), Robert Gurmankin (4%), Brian McCloskey (4%), Roger Royse (2%), Saed Sayad (<1%)

OUTLINE

1. Democratizing access to cancer patient data. (0:03)
2. Opening up cancer data for patients and caregivers. (3:10)
3. Personalized medicine and its challenges. (10:52)
4. Personalized cancer treatment and data analysis. (17:48)
5. Personalized healthcare interfaces. (27:09)
6. Reconciling conflicting data from multiple cancer diagnostic tests. (33:04)
7. Analyzing gene expression data in cancer diagnosis. (38:34)
8. Improving cancer care through data analysis. (45:47)
9. Using AI to analyze cancer data and predict outcomes. (54:36)
10. Cancer treatment decision-making with a focus on quality of life. (57:14)

SUMMARY

- **Democratizing access to cancer patient data.** 0:03
    - Frank Nothaft explains how he learned that there's valuable information to help patients, but it's not easily accessible, and he wants to open up data to democratize access.
- **Opening up cancer data for patients and caregivers.** 3:10
    - Frank discusses how to turn cancer data into actionable insights for patients and caregivers.
    - Frank has spent over a decade working in oncology and genomics, including time at Berkeley and several startups.
    - Frank's current goal is to make it easier for patients to work with their diagnostic data and make empowered decisions about their care.
- **Personalized medicine and its challenges.** 10:52
    - Frank shares his family experience with cancer diagnosis and advocates for a patient-centric approach in personalized medicine.
    - Personalized medicine promises to tailor treatments to individual patients' goals and needs, but current gaps in diagnostic testing and access can lead to incorrect treatments for 50-80% of patients.
    - Frank highlights the gap between medical knowledge and patient care, particularly in oncology, where new treatments are rapidly emerging but

  physicians and patients struggle to keep up with the latest research and side effects.
  - Frank suggests that software solutions could help address this issue by providing easily accessible updates on medical evidence, clinical guidelines, and new trials, improving the dissemination of knowledge and quality of care.
- **Personalized cancer treatment and data analysis.** 17:48
  - Brad Power highlights the gap between real-world evidence and clinical trials, emphasizing the need for personalized treatment options that take into account individual patient characteristics and preferences.
  - Brian McCloskey raises the challenge of doctors not knowing how to use genetic data to drive personalized treatment options, despite patients being armed with this information.
  - Brian McCloskey and Rick Stanton discuss challenges in leveraging liquid biopsy data for cancer diagnosis and treatment, with a focus on interpreting complex biomarkers and addressing gaps in care for advanced patients.
  - Roger Royse discusses the challenge of navigating the vast amount of cancer research data and the potential for AI to help organize and summarize it.
  - Jeff Krolick agrees with Roger and highlights the need for a scalable solution to help clinicians and patients stay up-to-date on the latest literature.
- **Personalized healthcare interfaces.** 27:09
  - Jeff Krolick suggests hiring a system navigator to help patients communicate with physicians and advocate for their needs.
  - Frank agrees and notes that this role could help bridge the gap between physicians and patients, ensuring that patients receive the appropriate care and support.
  - Frank highlights the importance of personalized interfaces for diagnostic specialties like oncology, where multiple doctors and diagnostic specialists may be involved in a patient's care.
  - Frank notes that the current standard of care often falls short in providing timely and personalized feedback to patients, leading to confusion and frustration.
- **Reconciling conflicting data from multiple cancer diagnostic tests.** 33:04
  - Jeff Krolick suggests creating a new specialty role to bridge the gap between AI, patients, and physicians.
  - Conflicting results between RNA sequencing and IHC for PDL1 protein in cancer diagnosis.
- **Analyzing gene expression data in cancer diagnosis.** 38:34
  - Rick Stanton suggests requesting RNA Seq data to clarify Tempus's 100% expression of PD-L1 in prostate cancer patients.
  - Robert Gurmankin mentions the difficulty in resolving a conflict in results between Caris and BostonGene tests.
  - Oncologists and diagnostic companies can share data more efficiently, leading to better diagnostic rates.
- **Improving cancer care through data analysis.** 45:47
  - Frank Nothaft highlights the problem of inconsistent diagnosis descriptions among doctors and trials, leading to significant consequences for patient treatment.
  - Frank Nothaft also emphasizes the importance of making clinical information and knowledge bases accessible to patients, citing examples from digital health.
  - Frank Nothaft is working on a project to make medical data more usable for patients, with the goal of providing actionable insights and helping patients get the most out of their data.

  - Frank Nothaft is seeking feedback from patients on the most impactful problems to address in their prototype, including how to convince clinical teams of their concerns and questions.
  - Brad Power suggests starting an online discussion forum to engage with a larger audience and continue the conversation beyond the live studio audience.
  - Brian McCloskey highlights the importance of longitudinal data analysis to understand the evolution of cancer and the effects of multiple rounds of therapy.
- **Using AI to analyze cancer data and predict outcomes.** 54:36
  - Brad Power and Frank discuss the importance of modeling in cancer research, particularly in understanding the progression of cancer over time.
  - Saed Sayad, a researcher from Toronto, is introduced as a potential connection for further collaboration in this area.
- **Cancer treatment decision-making with a focus on quality of life.** 57:14
  - Jeff Krolick wants more information on the quality of life side effects of cancer treatments, including real-world incidence rates and personalized decision-making.
  - Frank agrees, citing the importance of considering long-term management and day-to-day concerns for cancer patients.

“Opening up Access to Cancer Data for Patients" (Frank Nothaft) [#76]

TRANSCRIPT

Brad Power
We're happy to have Frank Nothaft with us.

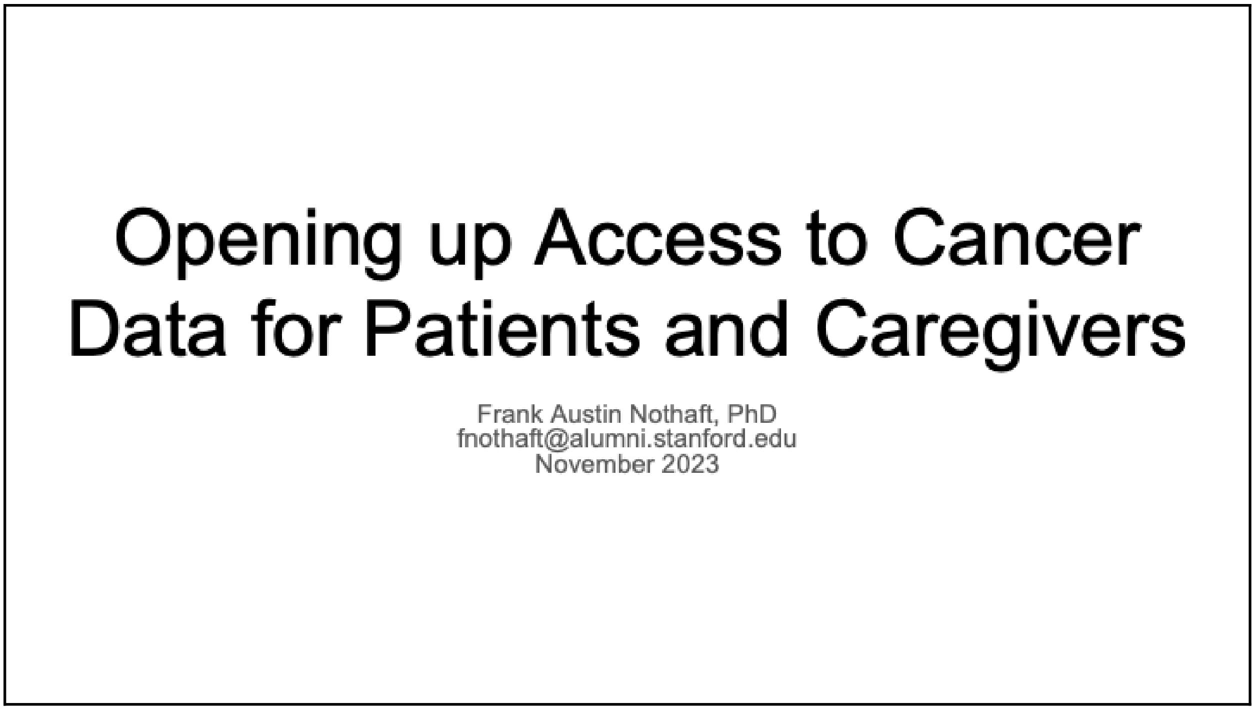
I connected with Frank when he was at Databricks, probably four plus years ago, when I was looking around at interesting companies who were doing innovative things when I was writing articles for the Harvard Business Review. Databricks was doing open source computation. They would do things like, if it took eight days to run a complex computer program, their software would make it run in four days. It was that level of complex computing. As consumers, we do everything in seconds, but they were dealing in days of computation. We kept in touch as he started working at Tempus. Tempus is a diagnostics company that many of us have used based in Chicago, and a friend – a company that is pushing the envelope in new tests and new information to help patients guide their care.

Frank left there a short while ago, and we were catching up. He explained that one of the things that he had learned is that there's a lot of information that could help patients, but it's not easily accessible to people other than those with his deep bioinformatics and research background. He wants to open up data and democratize its access so that more patients, like the activated patients in our community, can use it for guiding their care or, or if they're a caregiver for their loved ones.

This fits into a passion of mine, which is how we can disrupt healthcare by giving more information and power to patients.

Frank Nothaft 2:57
Thank you all for having me today.

“Opening up Access to Cancer Data for Patients" (Frank Nothaft) [#76]



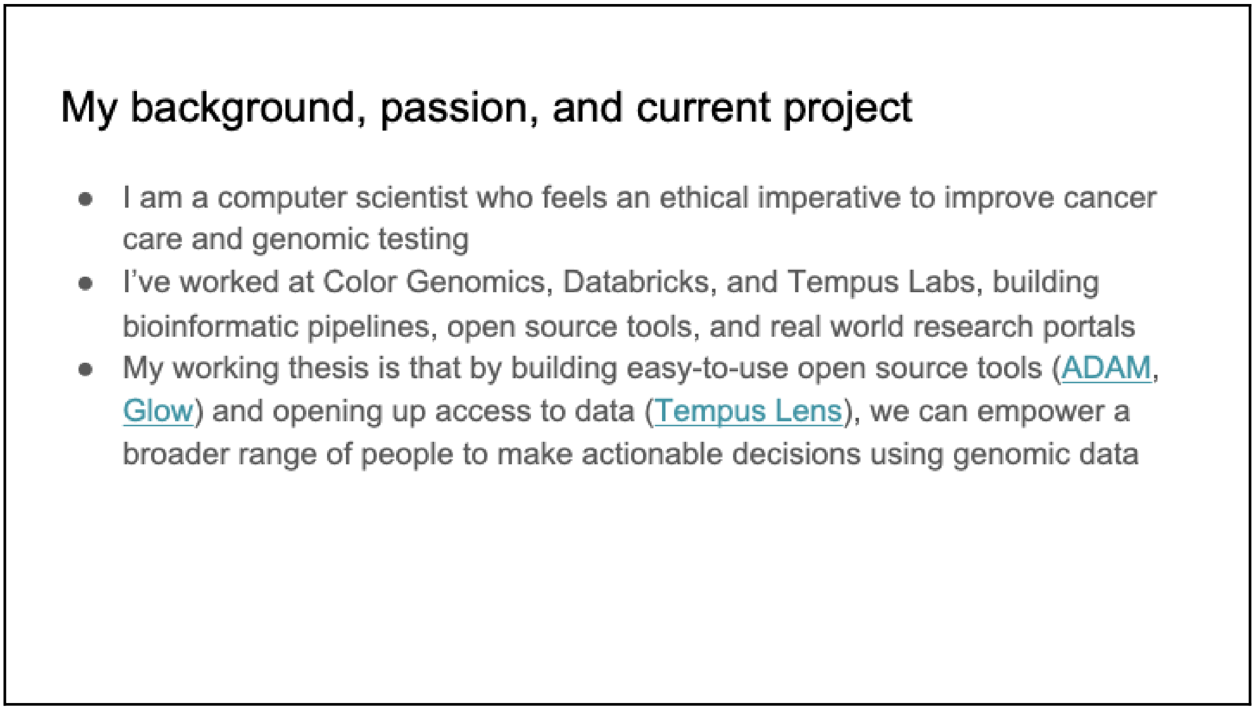
Opening up Access to Cancer Data for Patients and Caregivers

Frank Austin Nothaft, PhD
fnothaft@alumni.stanford.edu
November 2023

I titled this “Opening up access to cancer data for patients and caregivers.” Opening up data is a great ambition. But I think the real question is: “How do we turn data from just pure information into actual, actionable insight that we can use to drive our care in a way that is well aligned with our goals for the condition that we have?” That’s a very amorphous question. It's one that I've been thinking about and struggling with a bit recently.

I'd really love to get feedback from you all on areas that you think could be really helpful areas where you think data today is hard to use to drive decisions, decisions that you feel maybe slightly unempowered to make, or like you don't have good information and good data to support you in deciding.

“Opening up Access to Cancer Data for Patients" (Frank Nothaft) [#76]



My background, passion, and current project

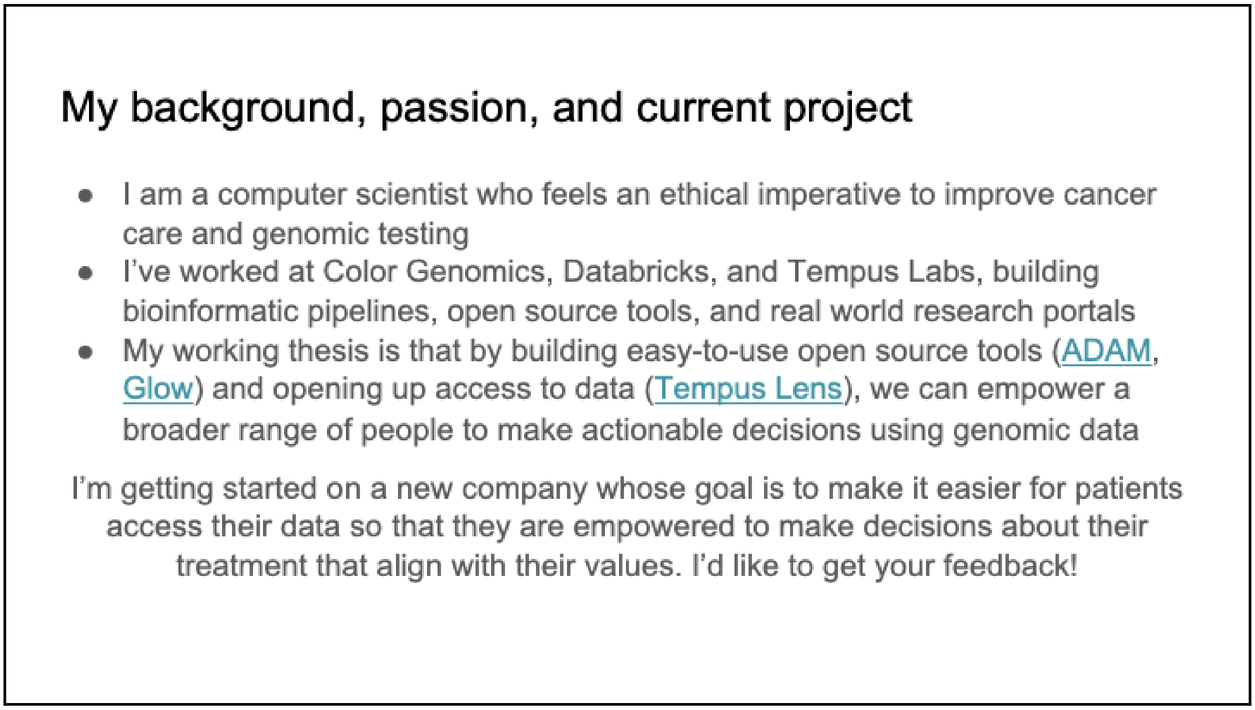
- I am a computer scientist who feels an ethical imperative to improve cancer care and genomic testing
- I’ve worked at Color Genomics, Databricks, and Tempus Labs, building bioinformatic pipelines, open source tools, and real world research portals
- My working thesis is that by building easy-to-use open source tools (ADAM, Glow) and opening up access to data (Tempus Lens), we can empower a broader range of people to make actionable decisions using genomic data

I'll start by talking a bit about myself, and I'll talk a little bit about what led me to this next phase of my life. I'm originally an electrical engineer. A little bit over a decade ago, I was talking with computer scientist Dave Patterson at UC Berkeley. He put out this really thought provoking article talking about how computer scientists might have the best skill set to make an impact in cancer care. Don't we, as ethical human beings, have a responsibility to pursue that? That really resonated with me. Over the last 10 years, I've spent my time working across diagnostic testing and research in both oncology and genomics, across academia, where I was at Berkeley, and then a couple of different startups:

- Color Genomics, where I worked on some of their very early hereditary cancer risk tests.
- Databricks, where we did a lot to help industrial organizations, both pharma companies, and diagnostic companies do large scale disease research and testing.
- Tempus, supporting a lot of our basic basic data research, basic and translational data research, worked on a platform called Tempus Lens that tried to make the data Tempus accessible at a true population level, to any qualified researcher that we were working with.

For me over the last decade, there's been one working thesis that I've been repeatedly hitting on: open source software. Open data is a really powerful lever that we can use to make this really powerful, really high impact data accessible to a much broader range of people who have important decisions that they want to make, whether it's for their care, whether it's for research into understanding a mechanism of what drives cancer, whether it's research to actually translate some of that basic insight into a new therapy in the clinic.

“Opening up Access to Cancer Data for Patients" (Frank Nothaft) [#76]



My background, passion, and current project

- I am a computer scientist who feels an ethical imperative to improve cancer care and genomic testing
- I’ve worked at Color Genomics, Databricks, and Tempus Labs, building bioinformatic pipelines, open source tools, and real world research portals
- My working thesis is that by building easy-to-use open source tools (ADAM, Glow) and opening up access to data (Tempus Lens), we can empower a broader range of people to make actionable decisions using genomic data

I’m getting started on a new company whose goal is to make it easier for patients access their data so that they are empowered to make decisions about their treatment that align with their values. I’d like to get your feedback!

In my next phase, my goal is to make it easier for patients to work with their diagnostic data and their EHR data and make empowered decisions about their care.

Something that's led me here is a couple of things in my personal life. About three years ago, my daughter, then my wife and I, were all diagnosed with autism. Autism is an interesting condition because a lot of the research and a lot of the clinical practice is actually driven not at all by patients. It's driven in a way that some patients see very controversially not aligned with their interests, what their needs are, and what the values that they espouse are. For us this was a really interesting time, since both I and my wife have a medical background, and we found ramping up on the literature to be a really big challenge for us. We spent a lot of time making heads and tails of the data. We're really lucky that we have that background. But for somebody who doesn't have that background, it's not easy to understand where the research is coming from, the biases in it, and get misled and do things that aren't necessarily aligned with your values as a patient.

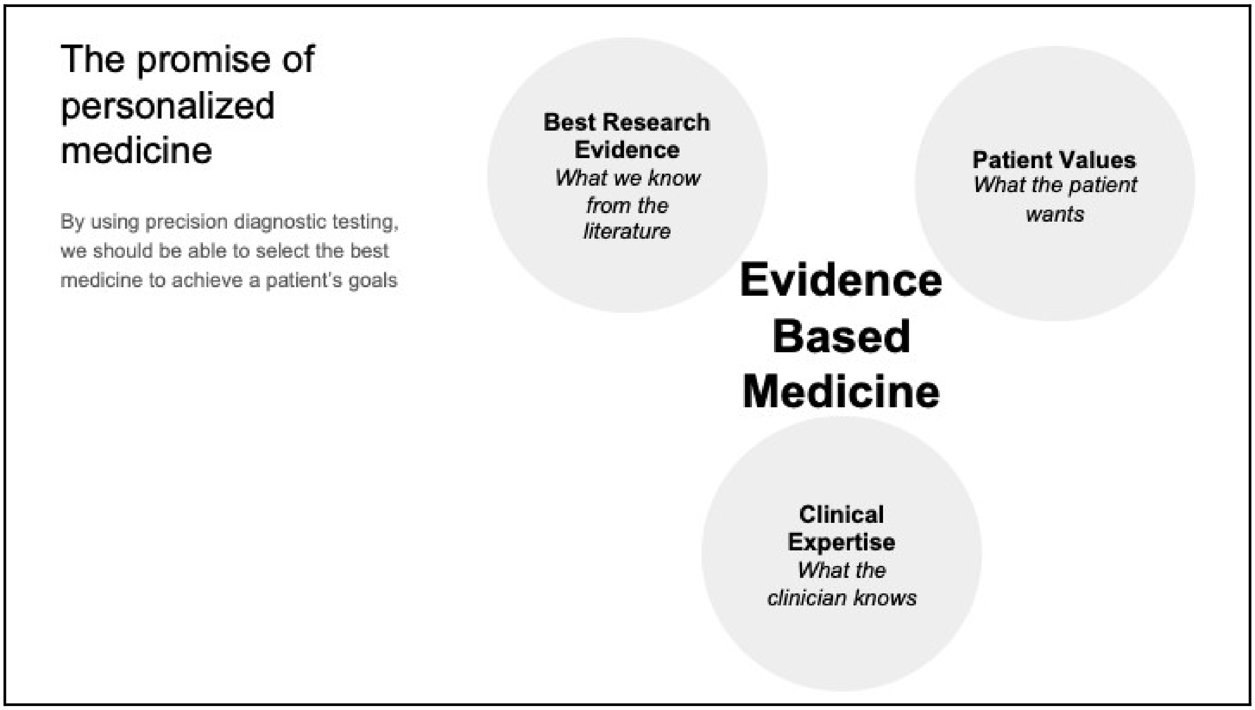
For me, as somebody who's worked in the diagnostic space, I've increasingly found that I am that person who, when someone close to me has a question about a diagnostic, they'll call me, and they'll ask questions. Over time, I found I feel really privileged. I've been able to help people in my life who have those questions.

A person really close to me has three sisters who have had breast or ovarian cancer. They were looking at getting diagnostic testing in that space to understand if they had a BRCA variant that would put them at risk. They had a lot of questions that they weren't able to really get answered by their doctors, given the time they were spending with them. I had this realization, that there are probably many patients, probably 99% of patients, who have gaps like that, where they have questions, and they're struggling to get answers.

It feels like something that we as a field, those of us who've worked in diagnostics, need to take a better patient-centric perspective and build things that really empower patients to center their values and their needs in their care.

That's a little bit about me and what's led me down this path.

I'll talk a little bit about some of my experiences in the personalized medicine space, and what I see as both the promises and the perils. I'd really love feedback on this because this represents my findings, my perspectives. I’m looking to generalize this and get other people's perspectives brought in.

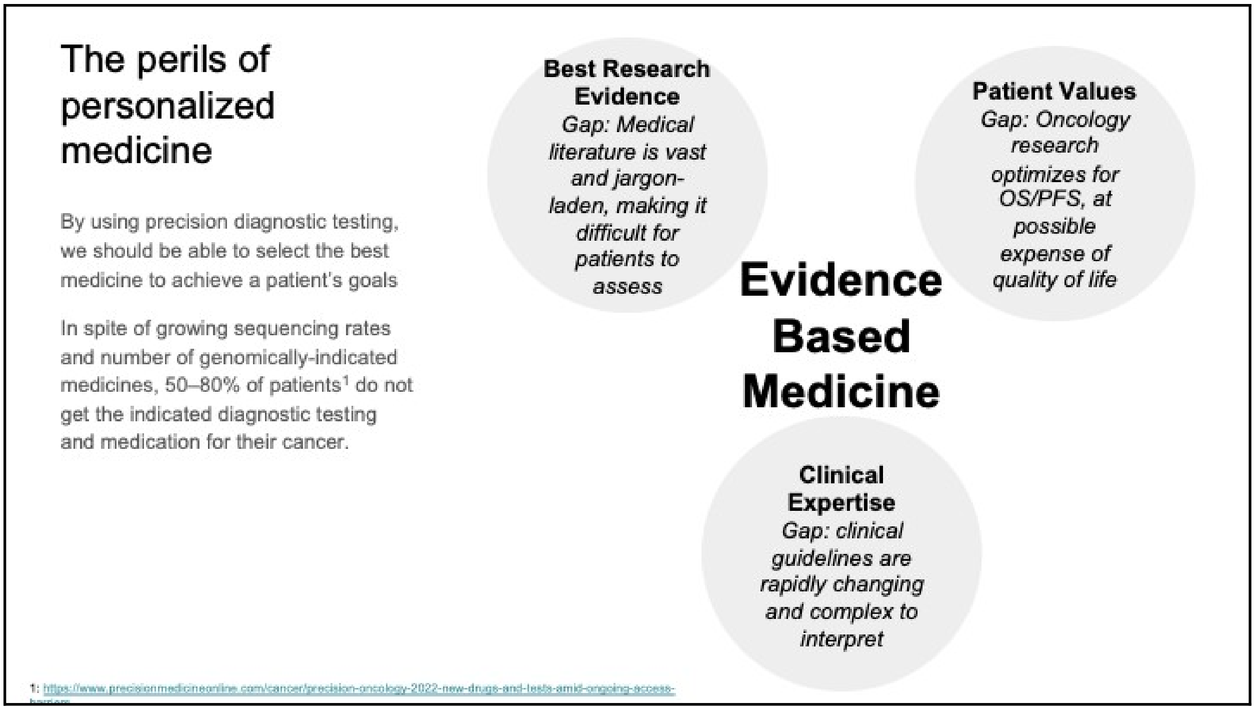


The promise of personalized medicine

By using precision diagnostic testing, we should be able to select the best medicine to achieve a patient's goals

One of the things that's really stood out to me as the promise of personalized medicine is that we're able to take the evidence that we know from the literature, that combines the research evidence that we have, all of that both basic clinical and translational research that we've done, and people have published about, that we have a good understanding of, and blend that with what the patient is looking to get out of their care and what the clinician has seen in the clinic. And hopefully, using all of these precision diagnostics, these next generation sequencing tests, new pathology algorithms that we see in the lab, to get to a point where we are picking a treatment that is most likely to meet the patient's goals for their treatment in an effective and safe manner. That's, in my mind, the promise of it.

“Opening up Access to Cancer Data for Patients" (Frank Nothaft) [#76]



The perils of personalized medicine

By using precision diagnostic testing, we should be able to select the best medicine to achieve a patient’s goals

In spite of growing sequencing rates and number of genomically-indicated medicines, 50–80% of patients[1] do not get the indicated diagnostic testing and medication for their cancer.

Best Research Evidence
Gap: Medical literature is vast and jargon-laden, making it difficult for patients to assess

Patient Values
Gap: Oncology research optimizes for OS/PFS, at possible expense of quality of life

Evidence Based Medicine

Clinical Expertise
Gap: clinical guidelines are rapidly changing and complex to interpret

1: https://www.precisionmedicineonline.com/cancer/precision-oncology-2022-new-drugs-and-tests-amid-ongoing-access-

One of the challenges that we found, and especially speaking from an industrial perspective, what still surprises me, is that about 50% to 80% of patients wind up not getting the “correct medicine” for their condition.

This comes from a number of different gaps that we have, for instance, some of the time patients wind up not getting the correct set of diagnostic testing that might be indicated for their condition. Some of the time that's because the diagnostic testing is too expensive and thus winds up being inaccessible. Some of the time the diagnostic test is not easy to access at the hospital or through the clinician that they're going with. Some of the time, there are other factors that lead into this.

Even in the cases where the patient does get the correct diagnostics, we hit this very interesting gap where the standard of care is moving so quickly in oncology these days, with new medicines coming out, better understanding of the side effects of medicines which have gone into the market, then administered to more patients. We've seen how patients actually fair in the real world on those. It's hard for physicians to keep up with the literature.

It's equally hard for patients who are trying to make these decisions to understand the full spectrum of what diagnostics, what therapeutics, what other interventions might be eligible to them. If they go on a medicine, what are some of the side effects that they might have? How can they hopefully mitigate those side effects so that they get the best possible response to the medicine?

You also have an interesting side effect, that comes out of how we typically approve medicines these days. We focus really heavily on an index of overall survival and progression free survival as our endpoints in oncology when we run clinical trials. It makes a lot of sense. They are the classical endpoints. But some of the time that may lead to quality of life expense that isn't necessarily evaluated in the phase three trial. Sometimes that may lead to therapies that are incredibly economically expensive.

There winds up being a large number of factors that lead to this gap. Some of these fields, to me, are very readily addressable. Others of these are very hard, complex, systemic ones to address, but especially where we have knowledge that is being generated, that we're struggling to disseminate, like updates to the medical evidence, updates to clinical guidelines, information about new diagnostics, and new trials that come out. Those feel like a very readily solvable problem in the software space. I'd really love to get any feedback here. I'm sure that I've missed things. We would love to get any, any, any thoughts?

Brad Power 17:47
Everything here is great. I would also add personalization versus population-based. Most of it goes to the point you have about how research is designed. Research is designed to work at population levels, not at individual levels. What's best for a patient who doesn't fit the population criteria? Puts them outside of access to treatments? That level of personalization can break the traditional research paradigm. When you get down to drug combinations unique for each patient, then how do you aggregate them to come up with some general principles that allow clinicians to deal with it?

My other favorite issue is real world evidence versus clinical trials. Patients need to make decisions now. If I look at the patients in our group, they're interested in things that are going to be available now or in the next six months. If you tell me that there's going to be a randomized clinical trial that will be out in five to ten years, that's beyond their horizon. It's nice that it's in the lab. But it must be accessible to be useful to me. The needs of patients are different in terms of time. Therefore, I want to see right now patients like me who are taking this drug and what their experience is. This is another gap in the system because the only evidence that qualifies as evidence is peer-reviewed, clinical trial evidence. Real world evidence is given short attention.

Frank Nothaft 19:44
Definitely agree there. I feel particularly passionate about that last one, because that was a big problem that we were trying to work on at Tempus. There's increasing regulatory openness to that. You're seeing more and more places where it is used, whether in submissions. But it's definitely a big gap, especially in empowering people who are making treatment decisions to use that evidence.

Brian McCloskey 20:28
Even when patients are armed with data coming from sequencing their DNA or RNA, doctors sometimes just don't know what to do with it. Particularly as you get into RNA or whole exome sequencing, there are so many flags that come up for them to be able to use that information to drive personalized treatment options. There's a whole host of barriers to leveraging this data. We've definitely pursued many, many different angles. One of the things is just finding a doctor

who actually can understand it, believes it, interrogate it, and then act on it. That's a big problem.

Frank Nothaft 21:27
That's a very spot on call out. Especially as we move into increasingly complex biomarkers, like some of the complex gene fusions, things like HER2-low in breast cancer. Those require even further understanding to interpret correctly. There can be a combination of getting an oncologist as well as a pathologist who feels comfortable working with us. I'd love to dive into that more and deeper.

Brian McCloskey 22:04
We've done a lot of work with Josh Bell at Tempus. I've been walking the halls of Tempus for about six or seven years.

Rick Stanton 22:23
I'm curious what you feel is of interest to this group. I'm interested in your new company. We've heard a lot of stuff. What can you do for us? Even if it's out in the future a vision statement. I don't see that yet. I see problems, I see gaps. I'm wondering what hope or value you can bring to this group. We're advanced cancer patients, out of standard of care, love to interpret data, don't have a lot of data necessarily. Maybe liquid biopsies would be something that probably everyone could get.

I'd like you to tailor a little bit towards what you can do to help us, if you don't mind.

Frank Nothaft 23:54
I have a slide in a minute where I'll talk a bit about what I've been thinking about and some of the open areas. I'm still trying to understand whether or not it's valuable.

Roger Royse 24:15
I listened to a podcast this morning about artificial intelligence and cancer care, and where that's going. I was just down at the Society for Immunotherapy for Cancer conference in San Diego last week. One of the things that becomes apparent to me is that there is just so much data out there, and there's so much information. It just seems like it's going to take computer modeling just to organize it all. On the podcast, a doctor commented, “I can't possibly read all the papers and studies that come out.” But a computer can read it. Tell me if it's important and summarize it for me.

It feels to me like that's going to be a really big part of managing the data going forward, and personalized care as well.

Frank Nothaft 25:10
I definitely agree with that. That's been a big point that I have been thinking a lot about. Because if you think about PubMed, PubMed is an amazing asset. But unless you have a lot of time to separate the wheat from the chaff, as well as a good medical and statistical background

to understand bias that is present in a study, what is the level of evidence that the study is creating, it's really hard to work your way through. I have gotten lost in it. I've seen other people close to me get lost in it, candidly. It's been a big problem that we would hear from people when we were at Tempus. That is a big part of what could be on offer. I'll talk about this a bit in the next slide. One of the great things about being in an organization like Tempus is you have a great network of people who are subject matter experts in really fine areas of minutia, who are up-to-date on the latest and greatest literature, and they are your ChatGPT.

For the literature, figuring out a way to go ahead and scale that would be valuable. But it's a question of how accurate it needs to be, and how people want to interact with it. How would people actually use it in their day-to-day, whether they're a clinician or a patient? That's an area that I'm a little fuzzy on right now.

Jeff Krolick 27:09
I have a behavioral health background, primarily as an administrator. One of the things that we recognize is whether it's somebody with a life-affecting mental illness, you're going to see a physician, a psychiatrist, a therapist. Clients don't always know, or patients don't always know what they want, what to ask. There's also the differential power dynamic between the physician and a patient. I went through three oncologists before I found the one that I'm currently working with and have done for quite a while. In this group, we're all very capable of doing that for ourselves. Most patients are not. One of the strategies that we use in behavioral health is to have a “system navigator”, who can help people talk to physicians and advocate. What I'm wondering, in this instance, is with physicians being as busy as they are, and as often I'll tell my primary care doc what I want and what I need, and I'll get it, but there won't be much discussion. It is almost like an interpretation specialist who understands how to use this and can translate it, can do that synthesis that you were talking about in your slide with the three different areas, and can talk to patients and can talk to physicians and help digest. Here's what we see from this.

That's my suggestion. We see it work pretty well in behavioral health. It helps people understand what they want, and to some extent, get what they want. This is a different level of detail and specialization, but it seems like there's almost a niche for that medical role to span those three areas and not just say, “The physicians can get it. They can click on this. They can get the summary.” Because they may not even know what the summary means or the implications of it, if they have enough time to do that.

Frank Nothaft 30:13
It can be useful in all diagnostic specialties of medicine. Oncology is a particularly relevant one, because it's very common that even if a patient is working with a single oncologist, there's a large care team. There may be multiple doctors, especially diagnostic specialists, like a radiologist or pathologist, who are supporting the care of the patient. There was an example that I had run across, not in oncology, but in another medical specialty, where someone close to me was working with a cardiologist and another diagnostic specialist, and the two specialists were looking at this differently with different diagnostic modalities. One was getting a very good very positive readout, the other was getting a very negative, very concerning readout. They were

struggling with how to go ahead and interpret this, and it led in a very adverse way. It's particularly challenging, too, because our standard for this in patient care and routing is to send your doctor a message in MyChart. When I've had concerns or questions about a readout that I've gotten, my recourse has been, “Oh, send the diagnostician, a message through MyChart.” They'll get back to you next week, which is really challenging as a patient. It doesn't give you that interactivity that you often need to really get to the bottom of something. It can feel like a really big mismatch between your urgency and the urgency that the rest of the healthcare system is giving you.

Brad Power 32:38
Jeff: Can you clarify what you're getting at? Are you saying that the interface to the information has to be thought about and that there has to be in the same way personalization of the treatment recommendations for the information that's coming to the patient? There also has to be personalization in the interface?

Jeff Krolick 33:04
From my experience, there are these positions that I would call “spanners”, who have a foot in different areas. You might say, “case management”, but that's not what I'm talking about. It's somebody that has enough information about each area, and can make those connections. In the medical field that's often up to a nurse to do that. Some nurses are very skilled, and some nurses are not. The quality of that isn't necessarily good.

I'm just wondering if, again, if there's a specialist role, that can bridge that gap between the patient, the AI readout of the best options, and the physician who likely doesn't have time even to sit at a computer and put all the data in?

There's also some concern about the black box. What's the quality of this information?

What I'm getting at is the creation of an almost specialty role in this at least at this point.

In behavioral health it has always been shocking to me that it's about a 15- to 20-year period of time from a well-established, well-researched intervention to it being commonly used in routine community mental health settings. We may be speeding things up here with AI.

Robert Gurmankin 35:24
There's a lot of data out there that I've gotten. But you have a lot of new companies coming on board. What happens when you get conflicting data from different companies? How do you reconcile it?

In my case, I looked at PD-L1, which is a common marker. It came back negative by immunohistochemistry, from a company called Caris. BostonGene did RNA sequencing and it came back 100%. So it's like,”Okay, your RNA is ready to make 100%.” Yet, the actual protein is not being seen at all, and it is needed. BostonGene couldn't rectify that discrepancy. I'm sure that if you get multiple reports, you're going to find multiple discrepancies because everybody's assays are not the same. Everybody's interpretations are not the same.

What's the patient supposed to do with it?

Frank Nothaft 36:46
It's a really good question, and candidly, a really tough one.

We did a number of different looks at that in our time at Tempus. Some of the time, it comes down to figuring out what is the most trustworthy test for a diagnostic purpose. For instance, for something like PD-L1, a cell surface protein, that might be a place where IHC would be a better read out than RNA seq. I'm not a PD-L1 expert, although I've done some research in the space. Take that as evocative, and not a specific recommendation. It is a little bit of coming up with guidelines for how to interpret those conflicting things. But honestly, you do see it in people who go through concurrent NGS testing, some of the time that is just that things are within the margin of error, especially if you're looking at low variant allele frequency mutations.

It's a really good question that I don't have a great answer for.

Robert Gurmankin 38:18
Fair enough.

Frank Nothaft 38:20
It's honestly just a very, very hard one. I realize that's not a very helpful answer.

Rick Stanton 38:34
I agree with Frank. IHC is generally regarded as more reliable. It's weird that you see 100%. Just to clarify, that would mean in RNA seq land that you're a high expressor. 100% compared to Tempus’s cohort. Out of let's say 1000 prostate cancer patients, that 100% might mean you're highly expressing PD-L1 compared to those patients. That's one, I would think that you would get from that. But the reason I raised my hand to chirp in here is that that's a point of 20,000 genes that Tempus is counting the expression of. So if I were you, I would ask to get the RNA seq file. It's just a little CSV file of the 20,000 genes, from Josh, and say, “Okay, does this make sense for the rest of my data?” In other words, you could do immune deconvolution and just see, “Do I have T-cells in my tumor microenvironment? Do I have CD-4 or CD-8 T-regs? What is the immune deconvolution of the 20,000?” You only need like 600 to make that the convolution. In other words, query that RNA seq data a little deeper. You have the data, or you can get the data that would help clarify it.

Frank Nothaft 40:42
To echo on top of Rick's comment, especially with things like RNA seq, it gets particularly challenging to interpret because oftentimes it's interpreted, as Rick was saying, to the distribution of patients in that cancer subtype. That can be the right approach in some cases. In some other cases, you would actually want to compare to a panel of “healthy patients” that have

been sampled from your tissue type, which is challenging. Sometimes we will look at an asset like the G Tech Data Set to do that sort of a comparison, because there can also be tissue type to tissue type biases that will cause gene expression to look a little bit wonky. That's most common in hepatic cancers. But I wouldn't be surprised if there are other other areas that it's a challenge for as well.

Rick Stanton 41:57
Did that help?

Robert Gurmankin 42:00
Yes. The PD-L1 immunohistochemistry came from Caris. I contacted them, and they're like, “Everything is in the report.” You're saying that they have more to offer, but it's more in a different form?

Rick Stanton 42:25
Where did you get the 100% from? Was that from Caris or from BostonGene?

Robert Gurmankin 42:29
BostonGene for the 100%. Meaning, compared to gene expression in their cohort of prostate cancer patients.

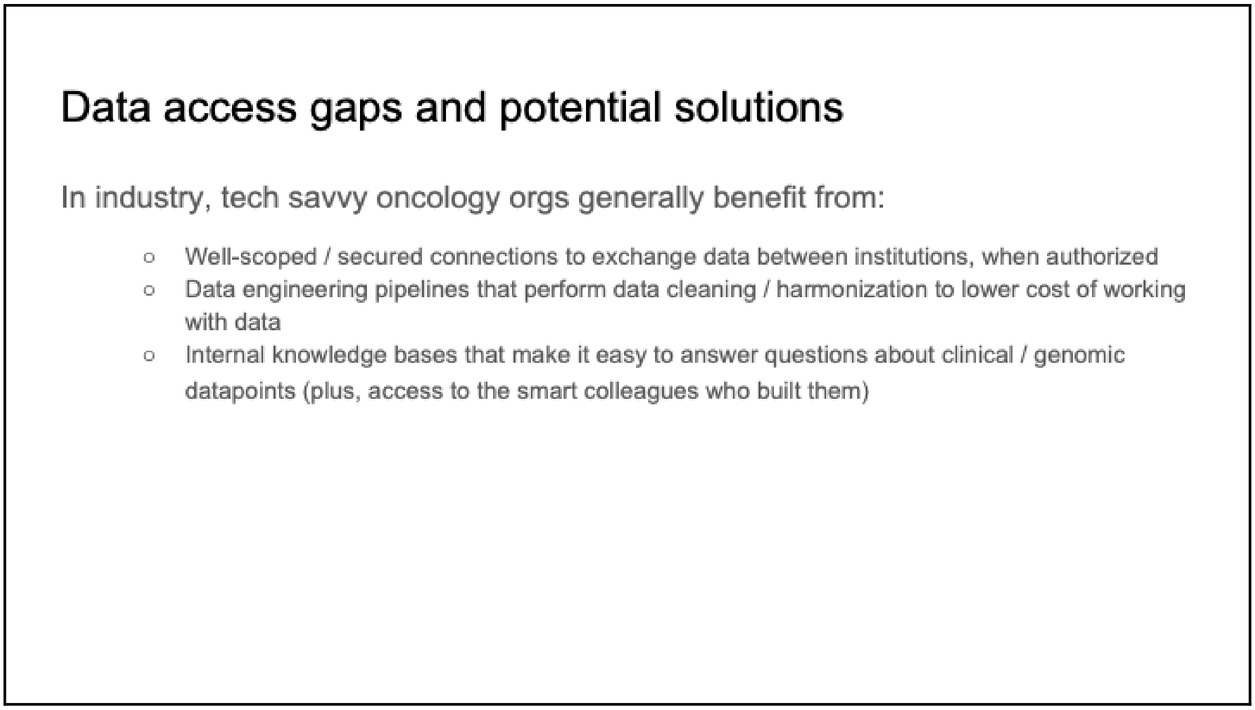
Rick Stanton 42:42
You have two different companies. You can look at the IHC. Is that possible? Would they be able to share the actual slide or the image?

Robert Gurmankin 42:57
I don't know. I can go back to Caris and see if I can get some raw data.

Rick Stanton 43:03
It's pretty easy to look at. Typically it's a brown stain. It's actually a little bit trickier because there are two different kinds of stains.

Robert Gurmankin 43:32
Maybe we can talk offline about how I get the raw data out of Caris.

“Opening up Access to Cancer Data for Patients" (Frank Nothaft) [#76]



Data access gaps and potential solutions

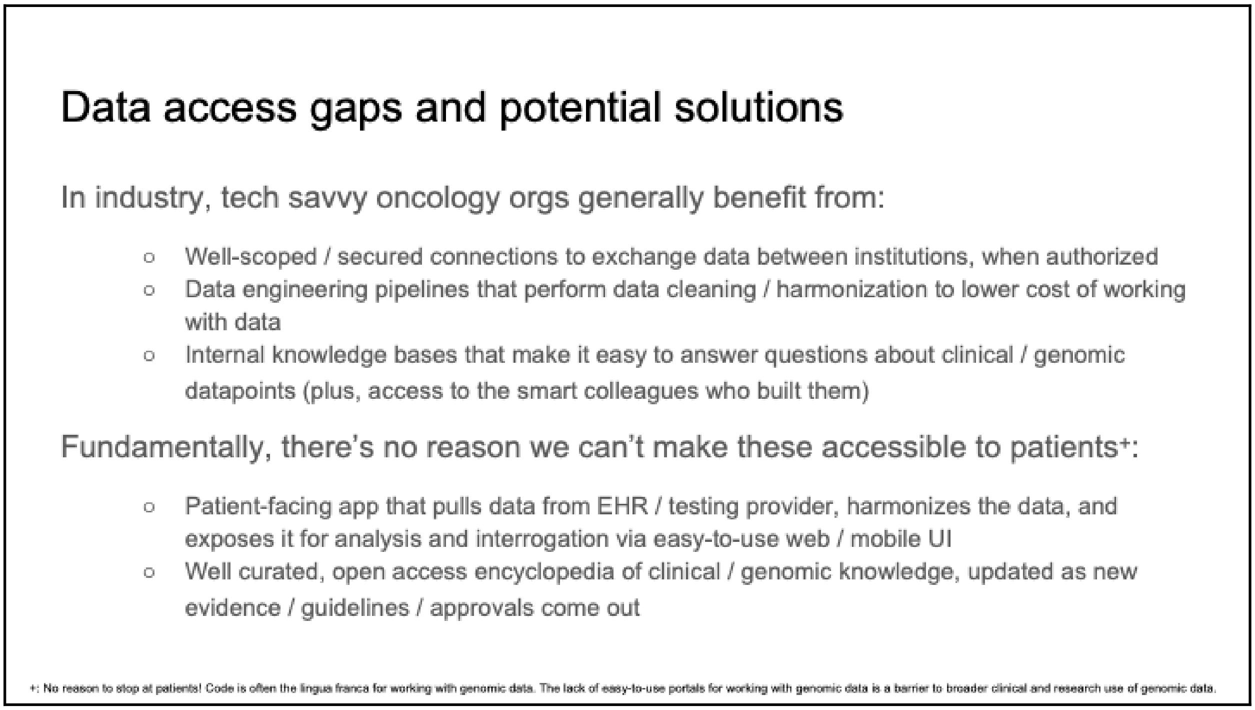
In industry, tech savvy oncology orgs generally benefit from:

- Well-scoped / secured connections to exchange data between institutions, when authorized
- Data engineering pipelines that perform data cleaning / harmonization to lower cost of working with data
- Internal knowledge bases that make it easy to answer questions about clinical / genomic datapoints (plus, access to the smart colleagues who built them)

Frank Nothaft 43:42
This is a good segue off of that last point. If you're in a tech savvy oncology org, whether it's a diagnostic company that has good technology buildouts like Tempus or Caris, or really, most diagnostic companies have pretty sophisticated tech stacks these days. Some hospital systems, and some clinical trial organizations, also have really sophisticated systems for this. When you look at the advantages that you have when you're sitting inside of one of these orgs, you typically have …:

- … really good, very complete, very well secured connections to get to share data between institutions, so the oncologist, pathologist, and the diagnostic testing company can share data in a way that allows them to get to a better diagnostic readout.
- … data engineering pipelines that perform a lot of useful data cleaning and harmonization. When I talk about lowering the cost of working with data, a great example is in the trial matching space. If you look at clinicaltrials.gov, they have a very unstructured way of representing inclusion and exclusion criteria, especially around diagnosis. Inside of a hospital or a diagnostics company, you might have a well-structured hierarchy for defining what a diagnosis is, and how diagnoses nest together. Tthis can be a really big problem in hematological cancers, as well as candidly in GI. There's a surprising amount of difference in how different doctors and different trials will describe similar diagnoses. This can mean very substantial things from the perspective of what medication someone is eligible to take. Tetting this cleaned up front can be really important.
- … very high access to internal knowledge bases. And some of the time this can be a medical team has gone through and done a great job of curating what are the approved therapies. In an easy-to-access way, what are the actual mutations, so on and so forth. Some of the time, we're just talking about gene expression. If someone has gone

through and curated, what do different levels of gene expression for this gene look like? Where are places that you have to be cautious in interpreting this? When you're inside one of these orgs, it's generally speaking, really easy to get access to the people who built these databases. They are oftentimes one Slack message away.



Data access gaps and potential solutions

In industry, tech savvy oncology orgs generally benefit from:

- Well-scoped / secured connections to exchange data between institutions, when authorized
- Data engineering pipelines that perform data cleaning / harmonization to lower cost of working with data
- Internal knowledge bases that make it easy to answer questions about clinical / genomic datapoints (plus, access to the smart colleagues who built them)

Fundamentally, there’s no reason we can’t make these accessible to patients+:

- Patient-facing app that pulls data from EHR / testing provider, harmonizes the data, and exposes it for analysis and interrogation via easy-to-use web / mobile UI
- Well curated, open access encyclopedia of clinical / genomic knowledge, updated as new evidence / guidelines / approvals come out

+: No reason to stop at patients! Code is often the lingua franca for working with genomic data. The lack of easy-to-use portals for working with genomic data is a barrier to broader clinical and research use of genomic data.

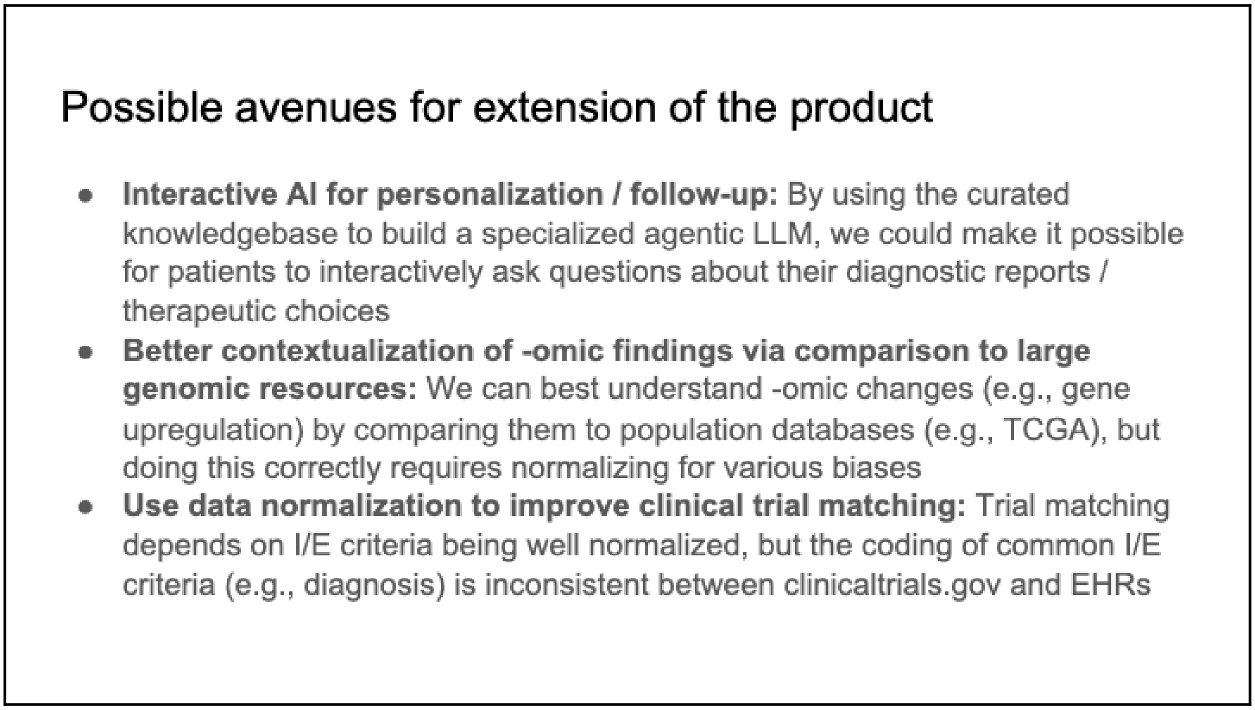
Fundamentally, there's no reason that these have to be proprietary. There's no reason that these can't be accessible to patients. We have lots of examples in the digital health space where someone has gone ahead and built a patient-facing application that pulls data from EHRs, pulls data from testing providers, harmonizes the data, and surfaces it. There are always regulations as to how you do that, but fundamentally it's a pattern that we've seen elsewhere in digital health.

If we think about curating a lot of that clinical information knowledge bases, like WebMD, and a lot of the cancer societies have done a pretty simple job of doing bits and pieces of that. But where I think it really does break down is going back to the fact that this data is very complex in cancer care and is getting increasingly increasingly subject to a patient's individual molecular profile. A lot of those knowledge bases are written to be very general, and balancing the complexity of making it fit a broad range of patients, while also making it very applicable to one individual who is looking for actionable insight for their care, given the information that they have, still remains a really big gap.

When you couple the two of these together, you should get to a system that makes the data a lot more usable, that hopefully exposes data that is clean and that has clinically good interpretations, and the right cautions on how to interpret it.

“Opening up Access to Cancer Data for Patients" (Frank Nothaft) [#76]

Ideally, that should allow us to build something on top where patients can really get into the meat of that data and ask the questions that they have at the time that they have it and get the insight that they need. That's the very broad frame of the idea that I'm working on.



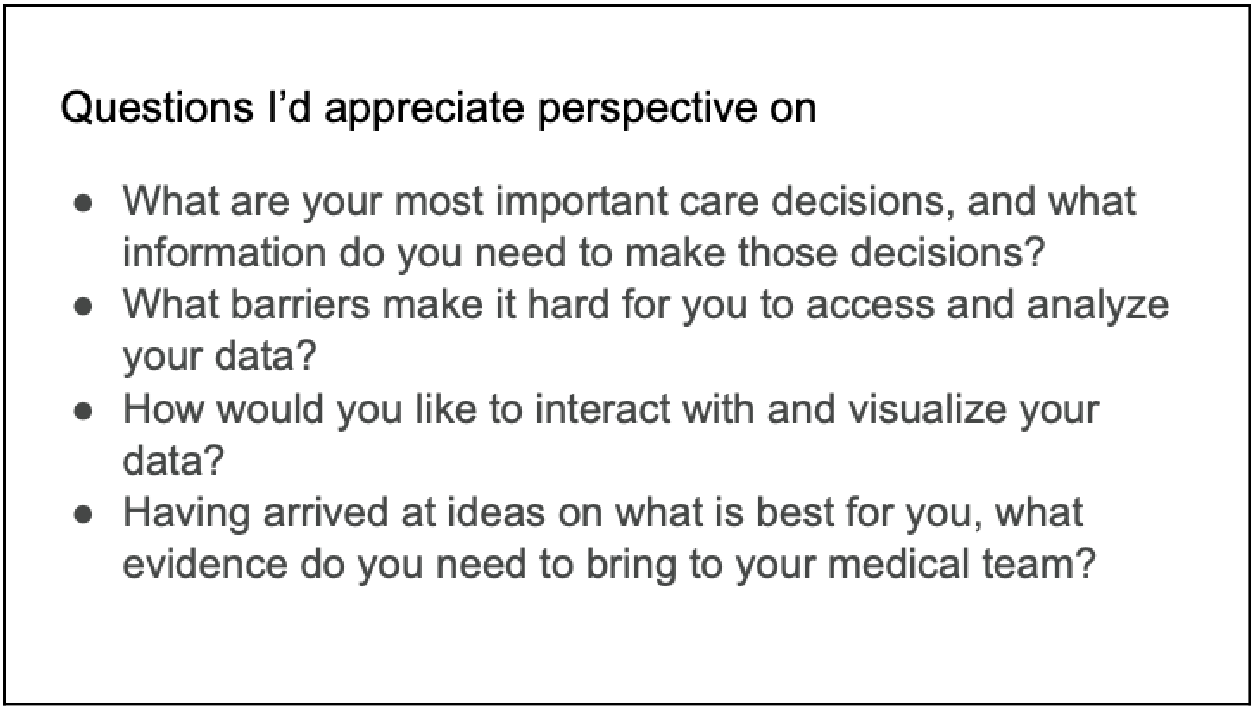
Possible avenues for extension of the product

- **Interactive AI for personalization / follow-up:** By using the curated knowledgebase to build a specialized agentic LLM, we could make it possible for patients to interactively ask questions about their diagnostic reports / therapeutic choices
- **Better contextualization of -omic findings via comparison to large genomic resources:** We can best understand -omic changes (e.g., gene upregulation) by comparing them to population databases (e.g., TCGA), but doing this correctly requires normalizing for various biases
- **Use data normalization to improve clinical trial matching:** Trial matching depends on I/E criteria being well normalized, but the coding of common I/E criteria (e.g., diagnosis) is inconsistent between clinicaltrials.gov and EHRs

Rick Stanton 49:50
We're big fans. Anything we can do to help?

Frank Nothaft 50:04
These are a couple of interesting avenues that we could go down to make this more helpful to patients, and we'd love to get feedback on.

To give you a sense, I'm starting to prototype this.



Questions I’d appreciate perspective on

- What are your most important care decisions, and what information do you need to make those decisions?
- What barriers make it hard for you to access and analyze your data?
- How would you like to interact with and visualize your data?
- Having arrived at ideas on what is best for you, what evidence do you need to bring to your medical team?

I would love to get really good patient-centered thoughts on like, “What are the most impactful problems to work on?” Here are some of the really big questions that I'm trying to dive into and understand. A really big one that has popped out today is the challenge of convincing the clinical teams that you're working with, about your concerns, about your questions. That sounds like a big problem.

Brad Power 51:32
We have a wonderful online discussion forum. In addition to your survey, you can start a thread that puts these questions out there, and we can have an interactive dialogue. We have about 18 people here today. This is our live studio audience. There are probably three or four times that many people in our community that would be happy to engage in a conversation online, but just couldn't make it at this time today. It's also continuous going forward. It can be people building on each other's comments. That gives us also a safety valve. We didn't get everything today. We can get to it going forward.

Brian McCloskey 52:36
Getting the most complete view of our cancer as possible, using complementary diagnostics, is a vision that I have. Also being able to tell the story longitudinally. Some of the prostate cancer patients on this call, who can live with this disease for a while, and I, go under many selective therapies, and that creates heterogeneity. Understanding where the cancer has been, and also predict where it's going, and what are the cause and effect of the treatments that we have, from a longitudinal perspective are really, really important.

Frank Nothaft 53:31

That definitely resonates, especially when you have multiple rounds of therapy, and when you have combination therapies. They're applying selective pressure on the tumor in a variety of different ways. That drives a response. It's a really big topic of research that we still have a fairly incomplete understanding of, and the only way to understand it is looking at the data longitudinally, which is very hard to do right now. That was a really big thing that we were working to solve at Tempus. Candidly, it is one of the biggest problems in oncology today, especially with the dominant clinical trial regime, where you are, for good reason, looking at things in a controlled environment, but that doesn't necessarily generalize to the real world.

Brad Power 54:36
Building on what Brian was saying, there are three categories here. There's the data. Then there's the AI that would interpret the data as some machine that can synthesize the data and come up with recommendations and keep on top of the data and all the things that IBM Watson was supposed to be able to do. But the third thing is disease modeling, as well. That's progression over time.

I probably mentioned to you, that we're big fans of Bob Gatenby, and the modeling that his team is doing at Moffitt Cancer Center. He's using evolutionary biology and game theory, thinking of the cancer as a heterogeneous set of clones, and as you're perturbing that system with treatments, what is it doing to the system? How is that then progressing?

The area of modeling is a third category of something that would be part of whatever the software architecture that you're building.

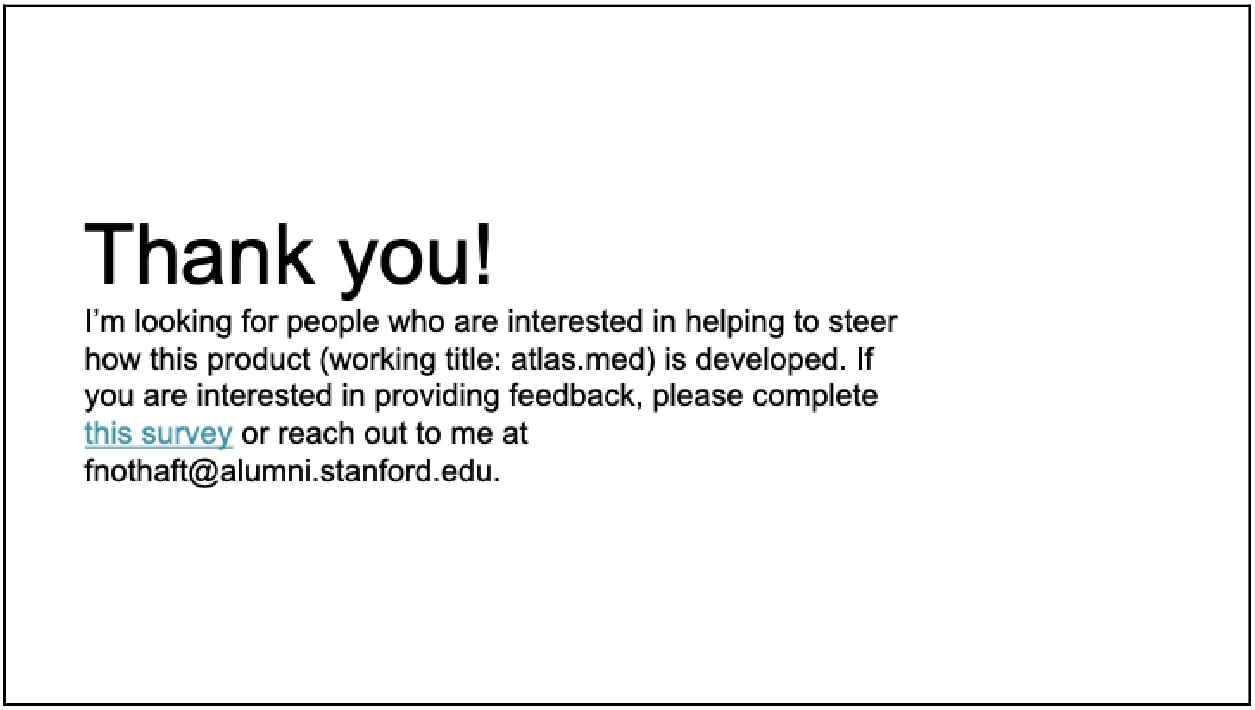
Frank Nothaft 55:39
That has been a pretty big gap. Both people in oncology, and increasingly people in germline human genetics, are becoming aware of this because having the data but not being able to make an actual statement as to what is likely to happen makes it just yet more abstract to reason over any of the data, which I think a lot of clinicians and a lot of patients really struggle with.

Brad Power 56:16
Saed Sayad of Bioada, who analyzes public databases, is a good contact for you.

Jeff Krolick 57:14
What I know would be helpful to me, and for others is, which is embedded in these questions, is a decisional balance. I'm looking more at quality of life, not just, “Here's the likely outcome, the percentages.” But, “Okay, if you try this new medication, here’s the real world incidence of these side effects you can look up on WebMD.” You can see two dozen things. Most people don't get these. But when you look at the clinical trials that you'd want to dig down, you'll see that 45% of people had this side effect, and level three or four are concerned with that. That information, at least in my experience, is never really discussed. And for me, that's an important aspect of making decisions , based on quality of life. Do I want this side effect? Can I live with that? Am I likely to get the side effect, not just the benefit of the treatment? But do I want those side effects

versus what I'm doing now, that's working pretty well, but maybe this will put me into longer term remission? So, for me, that's tying that all together in an information piece that would be very helpful for me.

Frank Nothaft 59:02
That makes a lot of sense. You have definitely heard that before from folks that are long-term managing their cancer, because, to your point, that's a very salient day-in, day-out concern.



Thank you!

I’m looking for people who are interested in helping to steer how this product (working title: atlas.med) is developed. If you are interested in providing feedback, please complete this survey or reach out to me at fnothaft@alumni.stanford.edu.

Frank Nothaft 59:33
Here's my survey link. I'll post it community.cancerpatientlab.org.

The working title for this is either “Atlas.net” or “Atlas.bio”, subject to change.

I really appreciate you all inviting me today, asking great questions, challenging me, and giving me feedback.