Eric Lefkofsky Blogs Compilation

Source: http://lefkofsky.com/

Credit: Hassan Bhatti

PUTTING WASTE TO WORK

About a dozen years ago, my wife and I set up a family foundation. Generally, we were focused on supporting four causes that we felt needed more attention: education, human rights, civic causes and medical discoveries. These days, I have become 100% focused on the last and I can't seem to think about anything else.

While that level of fixation is not healthy and while these other causes are beyond worthy, I feel the need to explain why I'm so singularly focused.

Let's start with money.

We spend about ~$3 trillion a year on health care in this country, and by almost all accounts (independent government studies, large Payor analyses, etc.) roughly one-third is wasted. That means we are spending roughly a trillion dollars a year that we don't need to be spending.

While that may seem like a manageable number since it's only twelve zeros, it's worth looking at the number in its entirety to fully understand its size:

$1,000,000,000,000

It's a big number.

More importantly, I believe that we are standing at the gateway of a new era of technology and medicine, one that is going to completely upend how we treat patients and manage disease.

I believe that by bringing big data (along with machine learning and artificial intelligence) to healthcare, we can reduce mortalities by well over 50% in the next 25 years and remove a significant portion of that trillion-dollar waste.

So, what would that mean for the myriad of other causes that we, like others, support?

Let's start with education. We spend ~$640 billion a year on education in this country. Our education budget has risen 3.5% over the past two years. That seems a bit light. Since we just saved a trillion dollars, let's say we were to increase the education budget by 20% a year (a staggering number that would most certainly improve our schools), that still leaves us $870 billion to spend.

Want to impact crime? We spend $80 billion on prisons a year. We could allocate another $80 billion (doubling the budget) towards working with individuals from disadvantaged communities, as well as providing job training and other opportunities to individuals who are already in the prison system, still leaving us with $790 billion to go.

How about tackling poverty. For ~$150 billion or so we can bring everyone in this country above the poverty line.

So what do we do with the $640 billion that's left. Maybe we want to retire our national debt or reserve for the interest we pay. If we chose the latter, that's $240 billion annually. So even after setting aside enough money to cover our national debt without using a penny of tax payer money, we still have $400 billion left.

And let's not forget we improved education, helped those living below the poverty line and worked on fixing our prison system along the way.

But wait, we forgot to accrue for the positive benefit of (1) an extra one million productive Americans who are still alive because they didn't die of cancer, or a heart attack, or a stroke and (2) a more educated society and (3) less people in jail and (4) no more poverty to contend with. Even if we assume conservative GDP lifts for the above, you probably pick up $500-600 billion annually, which means we're left trying to spend the extra trillion dollars we saved by improving our healthcare system and controlling diseases such as cancer.

In other words, remove the trillion dollars of waste and save lives with more effective prevention and treatment, and you have enough money to attack national problems with plenty to spare.

The elephant in the room is cancer (and other diseases) that endlessly consume our resources, impoverish our healthcare system, and deflate the spirit of every patient and family member battling disease.

And while up until now we were powerless to combat illnesses such as cancer, overwhelmed by its complexity, we now have for the first time, the tools we need to peer inside the body and understand what makes us healthy and what makes us sick.

Advances in molecular sequencing, machine learning, and artificial intelligence have armed us in ways that were truly unimaginable by those fighting on the front line.

It's time to double down. It's time to focus.

Because as you can see, if we win this battle, we win the entire war.

DEATH TO CANCER, BYTE BY BYTE

There has never been a more urgent time to harness and apply data science to help eradicate cancer. Why? For the first time, it is possible. We stand at the crossroads of two transformative advances in science. The first of these is our ability to sequence the entire genome of a cancer patient. The cost of genetic sequencing has fallen over one million-fold since 2000, making the technology finally accessible at scale. The second critical advancement is our growing appreciation and understanding of how to leverage the human immune system to fight cancer. One of the great challenges for our generation will be to invent a data-driven approach that can match a patient's genetic makeup to their optimal cancer therapy.

Recently, the FDA announced approval of an immunotherapy to treat cancers with a specific genomic profile. This marks the first time the agency has used a genomic profile rather than an anatomical tumor type (i.e. breast or pancreatic cancer) as a criterion in the drug approval process. However, it remains unclear why some patients respond well to immunotherapy, and others not well at all. At Tempus, we have assembled a team of roughly one hundred and fifty computational biologists, geneticists, and immunologists who, together with our data scientists and engineers, are pursuing an answer to this question: Why do some patients respond well to a cancer therapy while others do not?

We believe that combining the patient's genomic and clinical data is key to determining which cancer therapy should be implemented in a specific clinical context. We believe this represents the future of cancer care. To get there, however, we must apply many of the substantial advances made in computing over the past twenty years, from computer vision for analyzing tumor biopsy images, to natural language processing for parsing clinical records, to machine learning for integrating DNA and RNA information.

In the last century, every generation, emboldened by a disrupting technology, can lay claim to a monumental achievement that changed the way we live — from curing polio to landing a man on the moon. The challenge of our time is to end cancer as we know it. We finally have the tools.

I WONDER IF…

The other day I read a letter written by my partner, Brad, to everyone at Uptake. In it, he referenced the power of asking yourself the most basic of questions, one of which was: "I wonder, if…"

Normally, it would be the kind of thing I would read once and not spend a ton of time thinking about again. But for some reason, I can't seem to let this one go.

I'm not sure why, but the more I think about it, the more I feel like there is something profoundly powerful in those three simple words: I…wonder…if…

No matter what I tack on the end of those words, I'm instantly teleported into an imaginary realm; a free space without walls or borders.

I wonder if…we'll ever have politicians that worry more about the country then getting re-elected.

I wonder if…we'll ever live in a world without guns and violence.

I wonder if…kids will ever have the same chance of success, no matter where they grow up.

I wonder if…we'll send a man to Mars in my lifetime.

I wonder if…we'll cure cancer.

Wondering is powerful, as it gives rise to disruption and innovation. It's also the nucleus of imagination, which is the very fuel that makes the impossible possible.

Steve Jobs wondered if he could put a computer in your pocket. Jeff Bezos wondered if he could build a virtual mall on the internet that sold anything and delivered it to your doorstep. Elon Musk wondered if he could make cars for the masses that didn't need fuel.

These impactful ideas began as nothing more than a spark of imagination. Someone was sitting around saying to themselves: "I wonder if," and suddenly a question gave rise to an idea, an idea gave rise to a plan, and a plan gave rise to a product or a business.

Every once and a while, someone asks me how I've been able to start companies over the past 20 years that have gone on to become meaningful in some way.

My typical answer is that these businesses were all born out of Brad and I trying to solve real problems.

But, there's another answer that is equally true. I'm a daydreamer – I always have been. When something doesn't make sense to me, I often find myself lost in some imaginary world where the problem doesn't exist or I've figured out the answer.

At the oddest of times (in my car, in the shower, when I'm working out), I find myself dreaming, and in those dreams, I've constructed an entire world where the problem has been solved by a product, service or company that I've conjured up. Down to the smallest detail, I construct an entire ecosystem in my mind.

And at some point, the dream is so intoxicating that I'm compelled to do everything in my power to make the dream a reality.

I have to believe this same thing happens to entrepreneurs everywhere.

One of my favorite quotes is from George Bernard Shaw, who said, "Some men see things as they are and ask why. Others dream things that never were and ask why not."

Maybe the world would be a better place if we found ways to foster the imaginative spirit that's inside all of us.

Maybe we need more people dreaming and asking themselves "why not?"

More people saying "I wonder if…"

A LITTLE MORE EMPATHY

The other morning I was listening to "Hurt" by Christina Aguilera. It's a song about a daughter who loses her father and never got a chance to tell him how much he meant to her or how sorry she was that their relationship had deteriorated. There's a line in the chorus where she says "I'm sorry for blaming you for everything I just couldn't do, and I've hurt myself by hurting you."

As I was making a cup of coffee listening to that song blasting through my cordless headphones, I started getting all choked up. With tears welling in my eyes, I asked myself "why in god's name am I crying"? I'm not a daughter. My father isn't dead. We don't have an estranged relationship. In fact, I talk to him almost every day and the last time I hurt him was when I refused to stay home on a Saturday night in 1987 after I had my wisdom teeth pulled.
And then it hit me…empathy

On its surface, it seems basic.

The ability to see the other side, to imagine what it would be like to be someone else and feel their pain or have to deal with their lot in life.

But too often it is easier to see the world from our own vantage point, which can be limiting. It's true in personal relationships, it's true in politics and it's true in business.
In business, it's critical to see the other side. If you want to win a negotiation or convince someone to follow your lead or avoid confrontation, you need to understand where the other person is coming from. If not, you're blind to their perspective and blind to what motivates them. And business is all about motivation.
These days I spend a ton of time dealing with doctors and patients who are battling cancer. It's easier to have empathy, because the reality of the disease and what it means for the people diagnosed and their loved ones is so miserable that if it doesn't tug at your heart, you likely don't have one. The heartache is that palpable.
It's harder to have empathy in our everyday lives, but equally important. And while none of us are perfect (at least I'm certainly not), it seems like a hell of a New Year's resolution – to have more empathy.
Imagine a world full of more of it.
It would be nearly impossible to hurt someone if you could feel their pain. No murder, no rape, no assault. No child abuse. No bitter divorces. No bullies on the playground.
Wars would be nearly impossible to wage. Lawsuits would be few and far between.

Almost every ounce of pain we cause as a people would be muted by the fact that it's harder to cause pain when you feel it yourself. Like the world's biggest voodoo doll, every time someone went to inflict harm they would first feel the impact of that harm on themselves.

And in that world, I would imagine our priorities would shift. Instead of spending billions on tanks and tankers, most of our tax dollars would be directed to eradicating disease and suffering. We would invest in science and healthcare so fathers and mothers and sons and daughters wouldn't have to suffer when they hear words like "cancer".
So maybe this year our collective New Year's resolution should be "a little more empathy." It can't hurt. ☺

THE COSMOS CONNECTION

We live in a complex world. Technology is all around us, industries are evolving faster than ever, and the pace of innovation is almost daunting.

It feels a bit like Ferris Bueller's classic line "life moves pretty fast. If you don't stop and look around once in a while, you could miss it."

I think we all feel at times as if the world is spinning too fast, and the rules of the game have become too hard to follow, that it's harder than ever to succeed at whatever you're trying to do, especially when you're trying to do something new.

Almost two years ago, I began to immerse myself in trying to understand cancer. Even though I had spent the last 20 years deploying technology across many industries (printing, logistics, media, local commerce, manufacturing), this was an entirely new world for me.

Everything about it was new. I had to learn a new language as physicians rarely speak English when it comes to discussing patients. I had to learn entirely new fields of study, like biology and chemistry, at a level that would at least allow me to engage with the top minds in the field. I had to learn the nuances of giant industries, like healthcare and pharmaceuticals and insurance.

It hasn't been easy (as it never is) but the more you submerge yourself the easier it becomes over time. That said, there is one trick to learning something new that has helped me more than

anything else over time, it's basically my secret weapon: talk to anyone and everyone that can teach you something new.

At Tempus, even now, I spend an incredible amount of time talking to people – oncologists, pathologists, radiologists, surgeons, nurses, computational biologists, bioinformaticians, researchers, administrators, pharmacologists, insurance providers, and so on.

The more I talk to people and engage with others who have domain expertise, the more I absorb; a bit like osmosis.

And there is no better way to find people to connect with than to look for opportunities to collaborate.

One of the ways to encourage collaboration is to build and promote community. I've lived in Chicago for nearly two decades and we are known for a number of things: top universities, great museums and theater, incredible food, a vibrant downtown, the Cubs, and these days more and more a growing tech community (thanks to companies like Groupon, Grubhub and Uptake and incubators like 1871).

There are countless reasons that we should also be known for our contribution to healthcare, but we must first build a stronger and more innovative healthcare community in order for that to happen. We have all the ingredients needed for the ecosystem to thrive: world class medical schools, top ranked hospitals and physicians, large pharmaceutical companies in our backyard. The only thing missing is technology.

When we started InnerWorkings in 2001, Chicago didn't have much of a tech community, and too often our most talented technology grads flocked to the coasts. By 2008, when we started Groupon, the exodus of talent had begun to subside. By giving people a reason to stay (jobs, the prospect of more interesting projects and a sense of community) companies like Groupon helped create a community that is flourishing today.

This is precisely what the Chicago healthcare industry needs to do with computational scientists and other technologists who want to invest the time to learn what I have learned and help drive innovation in the field.

Technology has transformed every industry and every part of our lives, and its impact on healthcare in inevitable and will, I believe, have an unimaginable impact on our lives.

It's in our collective interest to build a health-tech community that is rooted right here in Chicago. That is the thought behind a Meetup group called COSMOS, comprised of computational biologists, bioinformaticians and those specializing in the analysis and/or computation of molecular data, to network and strengthen community. Too often, these men and women aren't integrated into the larger healthcare community and they should be.
If we hope to tackle a disease that has been around since the dawn of humanity, we need all of the best minds we can get collaborating and learning from each other. And we need to embrace technology and put our best software engineers and data scientists right next to doctors, working hand in hand to deliver the benefits of technology to the patient's they're serving.

It might seem daunting, but the first step is manageable – just reach out to someone, ask them a question, and learn something new. It's how some of the best collaborations, and partnerships, begin.

AN OPERATING SYSTEM FOR CANCER

By Richard Sallari, Kevin White & Eric Lefkofsky

Introduction

Since the initial draft of the human genome was released in 2000, we have collected an unprecedented amount of genomic data as a community. Comprehensive views of cancer genomes were first produced in 2007. For the past decade, a great deal of our scientific energy, as it relates to mining this data, has been focused on the hunt for driver genes that can cause cancer, known as oncogenes and tumor suppressors, and using this biological knowledge to develop new targeted therapies and therapeutic strategies. But perhaps the most important lesson we have learned so far is that we will need much more data in order to fully understand cancer.

Nixon declared war on cancer in 1971. For the past 45 years, we have invested hundreds of billions of dollars in understanding the drivers behind cancer. The government alone has invested over $90 billion since 1995 [1]. After 2000 our pace of identifying oncogenes and tumor suppressors accelerated, and yet today we still have only identified around 250 cancer driver genes [2], and only about 27 [3] are well established clinically and directly tied to FDA approved drugs. To put that in perspective, there are over 20,000 genes and there is an order of magnitude more gene-controlling DNA sequence in the portion of the genome that does not encode proteins (also referred to as "non-coding"); so after all this time and money, we still have only mapped a small percentage of the genome as it relates to oncology. Although it has been argued that a plateau has been reached with regard to identifying the most common cancer driver genes [4] we have a long road ahead before we have a comprehensive map of all segments of the genome that are relevant to cancer. And we have barely started mapping genes that drive metastases or the associations between genes and therapies in a formal and systematic manner, in large part because the process requires vast amounts of data that up until now have been too expensive to collect. The comprehensive search for driver genes in primary tumors alone is estimated to require 100,000 patients [2]. If we also consider metastasis and therapeutic associations, a mapping of cancer biology and patient response will require the joint analysis of even more patients.

The requirement for large data sets is due in part to the heterogeneity of cancer. Each patient's tumor is, to a great extent, unique; each cancer a novel manifestation of the relentless force of evolution thrust against the individual. Cancer is also heterogeneous within each patient. A tumor is often made up of many different cell populations, some capable of resisting treatment and others able to metastasize to distant locations; and these mutant cells often cohabitate within the same tumor geometry. And yet today, the largest genomic data sets we have amassed are through The Cancer Genome Atlas (TCGA) and International Cancer Genome Consortium (ICGC) and cover roughly 15 thousand unique exomes (the 1.5% of the genome that codes for protein) [6]. Unfortunately, the difficulty of cancer biology makes this data inadequate if we are to envision an end to cancer in our lifetimes; the volume of patients required is far beyond what any single hospital or institution can collect or analyze. The complexity of the cancer problem requires a highly scalable technology solution.

An Institutional Problem

The comprehensive elucidation of cancer biology amounts to a massive search algorithm executed in parallel across hundreds of institutions globally. Universities, hospitals and companies are all probing the myriad facets of cancer biology in search of insight. Some researchers specialize in a given cancer subtype, like lung or breast, others tackle a specific gene or biomarker, while others put their faith in novel drug discovery. Most of these individual efforts are rational and intelligent, often brilliant, but at the macro scale the current approach represents a brute force implementation, where each institution is looking for a piece of the puzzle largely blind to other efforts.

The biology of cancer is difficult, as the highly redundant system that evolved to keep us alive transforms into the very enemy we face when fighting cancer. Our bodies have adapted to ensure that there are numerous ways to achieve critical functionality. Mutated and dysregulated genes that drive cancer are a part of that complex system; as such, the task of the massive search algorithm embodied in the basic, translation and clinical research of the global cancer community is to not only identify them all, but also to understand their relationship with the circuitry of each cell type, including how they affect and relate to the non-coding portion of the genome. And while identifying driver genes might seem relatively straightforward now, the combinatorial nature of cancer therapies (chemo, radio, targeted or immune) will likely require exponentially larger datasets to discover novel therapeutic associations. This is due, in part, to the curse of dimensionality where, as the degree of nuance in the patterns we search increases, the amount of data available to interrogate each of the patterns becomes sparser.

We are thus confronted with a statistical hurdle that cannot be easily overcome. The inefficiencies of the current macro strategy are frequently referred to in the community as “misaligned incentives” [7]. But the truth is that the cancer community, which is composed of patients, nurses, doctors, researchers, professors, legislators, lobbyists, insurers, advocates, and many other professionals is a complex system in its own right, with emergent behaviors that are not entirely predictable, or always operating in a rational manner. And because the system is riddled with regulation, and highly dependent on academic institutions and governmental agencies, the pace of innovation and adoption of new technologies has lagged behind other sectors.

An Industrial Solution

Aligning the cancer community might be as complex a task as curing cancer itself, but simply stating its flaws is an exercise in futility. Any effort to realign the system will require, at the very least, a platform for everyone involved to orient themselves around data (molecular, phenotypic, and therapeutic), that can be accessed and analyzed in both the clinical and research setting in a frictionless manner. What is needed is an industrial scale systems biology platform for cancer care, with state-of-the-art decision support tools for oncologists, pathologists, and surgeons. A platform to gather large amounts of molecular data, combine it with phenotypic and therapeutic data, analyze it in search of clinically relevant patterns, and ultimately test alternative therapies in patient-derived biological models. An operating system for cancer, if you will.

With the volume of clinical trials and new therapeutic options rising, especially those tied to molecular data, oncologists have the nearly impossible task of keeping up with ever-increasing amounts of data being generated during the course of a patient's treatment. This problem is aggravated by a lack of technology and infrastructure that physicians should be able to access. Most Electronic Medical Record (EMR) and Electronic Health Record (EHR) systems were designed decades ago and have failed to keep up with modern software and database architecture improvements. Data is highly siloed and often locked, analysis is cumbersome and requires significant time and bandwidth, and feature development and support tools are lacking. On average, physicians spend one to two hours on EHR and desk work for every hour of direct face time with patients. EHR burden is correlated with physician burnout, which is on the rise at a staggering 54% in 2014 [8]. This burden is likely to continue increasing, as most of the relevant patient data is actually in free text fields and is nearly unsearchable and unanalyzable without sophisticated Natural Language Processing (NLP) tools or advanced machine learning. As such, physicians need software and analytic tools that work within a hospital's existing infrastructure to analyze data and create a roadmap for patients who are not responding to the standard of care. Today, despite advances in genomic sequencing, physicians still lack the basic technology to analyze molecular data and use basic pattern recognition tools that are prevalent in other industries to help them deliver more effective and personalized care.

The technologies needed for such a system already exist. We can now sequence patients and gather rich amounts of molecular data at a fraction of the cost required just a decade ago [9]. We have the tools needed to bring deep learning and big data analytics to cancer, as the core infrastructure necessary to compile and analyze large data sets is within our reach. Additionally,

researchers have made incredible strides in the areas of Patient Derived Xenograft (PDX) modeling, organoid development, and microfluidic organs-on-chips such that we can build highly predictive *in vitro* and *in vivo* models of a patient's tumor. All that is needed now is for someone to invest in scaling these technologies and deploying them throughout hospitals worldwide.
Once the system is connected and data is aggregated, the problem becomes computational. How do we find patterns, associations, and connections among what appear to be unrelated or loosely related data sets? These problems are outside the area of expertise of biologists and chemists; but they are the everyday work of mathematicians and computer scientists.
As this new and budding operating system is constructed, marrying vast arrays of omic data with patient outcomes, we proffer that patterns will be recognized that will offer clinically relevant avenues for physicians and researchers to explore. Associations that have been historically invisible due to the limits of existing public and private databases will become recognizable with scale. Once discovered, validating these patterns will involve equal effort in the areas biological modeling and clinical trials. We recognize that there are numerous roadblocks and bottlenecks between discovery and clinical implementation. This is why we are deeply committed to both data-driven discovery and rigorous experimental validation.

A Biological Problem

The difficulty of cancer biology

The vast majority of cancer patients die from invasion of their critical tissues by metastatic cells. In contrast to many other diseases, these metastatic cells are not acquired from other patients, nor are they present in the patient at birth. Metastatic cells emerge and evolve from normal cells and do so independently from the cancer cells in any other patient; in each case, their trajectory is unique. With many other diseases we confront the end result of an evolutionary process. With cancer, we confront the evolutionary process itself. This is the source of difficulty for cancer biology. It is a disease of the genome, a product of molecular progression where apoptosis (programmed cell death) and mitosis (cell division) are hijacked, along with other hallmark molecular systems that in health allow us to grow, heal, and remain in balance [10]. As mutations occur in the genome, our molecular equilibrium is compromised, and cancer ensues.

Mutation is the driving force behind biological evolution. The accumulation of mutations in the genome of normal cells can eventually lead to the formation of a primary tumor. This mass of cells can be initially benign, simply expanding inside the body without harm to its host, but it can

also create an environment that fosters the exponential increase of mutations and variability in the cells that compose the growing tumor. The primary tumor is thus a springboard for more aggressive cells to evolve, as the environment is now ripe for cells to grow, propagate and survive in an uncontrolled manner. It is therefore important to understand the genes that, when mutated, drive the formation of primary tumors. However, it is even more critical to understand the genetic mutations and cellular states that lead to metastasis. Fortunately, despite every cancer being unique, there are some similarities between tumors. These are likely due to inherent genomic vulnerabilities, inherited variants that increase risk, environmental carcinogens, and convergent evolution. Mapping these similarities is critical to understanding the origins and progression of cancer.

The unfinished map of primary tumor drivers

The discovery of a single driver gene is usually the first step in the development of targeted or immune therapies. To date, despite considerable effort and vast sums of research investment, we have found about 250 statistically rigorous driver genes out of the pool of over 20,000 protein-coding genes. Through great feats of coordination, sequencing and computing, massive international research groups have defined a list of the most common cancer driver genes. However, they have also uncovered a “long tail” of recurrently mutated genes that also appear to bear the hallmarks of cancer drivers. Understanding how the common aberrations in these genes relate and interact with each other, and with the long tail of increasingly rare but recurrently mutated genes, is a major frontier facing cancer researchers today. This challenge requires large sample sizes and rigorous statistical models; methods that separate signal from noise in cancer genomes that can account for a variety of confounders, such as mutational heterogeneity across the genome. However, issues of statistical scale are the main hurdle if we are to identify the majority of cancer genes. The recent estimate of 100,000 patients across all cancer types will allow us to identify approximately 90% of driver genes affecting 2% or more of patients [2]. Current analyses for single tumor types are based on data sets of a few thousand exomes. These data sets are the result of nearly 10 years of cancer genome data accumulation. So how long will it take the community to collect ten times more data? In order to shorten the natural time horizon to break the statistical barrier of cancer genomics, we need to accelerate our sequencing efforts. For example, more than 60% of oncologists have never requested tumor sequencing in breast cancer [11]. Among those that do, the majority do so in less than 5% of

their patients; the primary limitations to use being access and funding [11]. Service providers that offer sequencing of even a few hundred genes are still prohibitively expensive and slow, despite significant hardware advancements that should have produced a greater impact on reducing prices. The first major challenge is so large it is beyond the scope of any single institution.

The uncharted landscape of metastasis drivers

While the discovery of the complete set of driver genes in primary tumors will take time, we have already found many of the frequent culprits for major cancer subtypes and at least there is a trajectory to accumulate the necessary data, albeit over a long period of time. The search for metastatic driver genes, however, is less developed. The vast majority of sequenced tumors in public databases are primary tumors that lack the additional transformative mutations that metastases acquire. Metastatic cells are the culmination of a long arms race with the host's immune system, often over several years. These cells have been shaped by each patient's unique battle against cancer; the drugs they took, their response, their environment, their diet, etc. They are therefore more heterogeneous and more evolved than primary tumor cells, and capable of adapting to a variety of therapeutic regimens. The mapping of all driver genes for metastasis may require an additional hundred thousand patients, if not many more. But for now, despite being responsible for the deaths of more than 90% of cancer patients, the metastatic space remains unmapped with virtually no drugs available for metastasis-specific gene alterations [12]. Many physicians continue to prescribe the same drugs to patients in an often linear fashion, as if the primary tumor was still the culprit years later. In many cases, it is not. The tumor has mutated in a manner that makes it largely disconnected from the initial disease. That said, just as we continue to find common driver genes in primary tumors, it is probably a matter of time before we identify common genomic culprits and cellular states that are driving metastasis. The same solutions that will speed up the process of data collection on the primary tumor side are necessary to help metastatic patients.

Bottlenecks in drug discovery

Side effects arise frequently when treating cancer because, by its very nature, the cancer we target is a part of us, and by "treating the disease" we often are poisoning ourselves. Survivors understand the hardships of untargeted chemotherapy and radiation treatment. In the cases where a targeted therapy exists, the complications and side effect profiles are lower by orders of

magnitude, because what is being treated is the thin margin which differentiates the tumor's biology from that of the patient. Identifying a new cancer gene is a significant scientific breakthrough, but it has little benefit for cancer patients unless it, or the mayhem it promulgates throughout the cell, can be targeted by a drug. Despite having identified dozens of cancer driver genes, only a small fraction of them are clinically actionable. As we perform systematic, scaled-up studies and identify even more cancer genes, we will have to design drugs for each of them. However, designing a drug is even harder than identifying a cancer gene in the first place, and it is getting harder every year. The cost of developing a new drug doubles approximately every nine years [13]. This problem is known as Eroom's law, which is a reverse spelling of Moore's law that described the doubling of components in integrated circuits every two years in the semiconductor industry. The ecosystem in which drugs are developed is routed in regulation, medicinal chemistry (which is as much a practitioner's art as it is a science), and a statistical methodology developed nearly a hundred years ago that relies on outcome data collected over a long period of time and validated through controlled, randomized trials. It is a system that made sense when we were largely testing highly toxic chemotherapies, but that must be re-examined in today's world of emerging targeted and immune therapies. When you take into consideration the number of chemical compounds implicated in oncology procedures (over 10,000) and their varying degrees of toxicity, pervasive institutional regulation and patient heterogeneity, it is no wonder that less than 200 cancer drugs have been approved by the FDA over the past 20 years [14].

Roadblocks in treatment data access

Unfortunately, identifying a cancer gene and developing a targeted drug for it are still insufficient to produce an effective and viable therapy. An additional layer of variability still needs to be overcome as every gene can be mutated in thousands of different ways. Additionally, every patient has their own genetic background, and every tumor might find a way to evade the drug's effect. In other words, even once you find a suitable target and a patient that initially responds, the cancer's genome can mutate again to produce new tumor cells that are resistant to whatever therapy was developed in the first place.

The mapping of all patient responses to therapy as a function of their genomes is largely uncharted territory. The confounders can be imagined, but remain untested, as we have no rigorous statistical framework with which to estimate the power of current approaches, let alone

to determine which new approaches would be necessary to obtain a comprehensive map. Because response is a combination of multiple genes and drugs, the combinatorics are much larger than for identifying single cancer drivers. Hence it is highly likely that the number of patients needed for such an endeavor will be even higher than those needed to identify all driver genes in both primary and metastatic tumors. The exploration of response and resistance is a daunting and ill-posed challenge. It requires something that currently does not exist at scale in any public or private data base, namely a massive data set that combines molecular, phenotypic and therapeutic data, that is also tied to patient derived biological modeling so we can boost power and rapidly analyze disease and response, and so that we can iterate in a time frame that mirrors the patient's own medical timeline. If a typical stage 4 metastatic patient has two years to live, we cannot take 10 years to test a small handful of therapeutic options; it is just too slow. We must build a system that is capable of real-time learning.

Outdated strategies in clinical trials

Mapping out the entire space of genes, drugs and responses is a massive undertaking that is beyond the scope of any single hospital or research institution. Given the current pace of collaboration, this could take decades to accomplish, but there is hope that disruptive technologies might be able to break through these statistical barriers and bring us to a solution much sooner. Therapies that, unlike a drug, can be tailored dynamically to the specific needs of a patient as their cancer evolves have great potential. T cells engineered to recognize tumor antigens, therapeutic CRISPR aimed directly at leveraging cancer mutations, or anti-metastatic drugs that prevent progression beyond a primary tumor are just a few examples. But again, these solutions require large amounts of data and a modern technology platform in order to be fully tested and implemented in a clinical setting, as they are a hostage to the same hurdles that all other therapeutics face today.

At the heart of pharmacological progress is the clinical trial. Clinical trials provide the basic evidence of efficacy and outcome that regulators, providers, and insurance companies rely on when they determine whether or not to approve a drug, prescribe a drug, or reimburse a patient for taking a drug. Trials are incredibly slow and expensive for a myriad of reasons, but perhaps the most important one that relates to our lack of rapid therapeutic progress is the mismatch between disease, trial design, and patient enrollment. Trials provide the evidence of efficacy of a drug within a specific clinical indication and are currently designed around cancer subtypes, for

example: women, pre-menopausal, who have stage I or II breast cancer, are node negative, and hormone positive. Clinical trial criteria regularly do not take into consideration molecular characteristics. In others words, the trial is not further limited to patients that fit the phenotype above and also have a PIK3CA and RB1 mutation. And yet, it is the very presence of these unique genomic elements that often governs which patients will and will not respond in the trial.

Trials are limited by cost, both of enrollment and validation. It is typically so expensive to enroll patients, and so time consuming, that drug companies often prefer to design trials that are broadly defined. The cost of conducting a phase I trial (where a drug is first screened in humans; primarily concerned with safety and tolerability) can range between $40,000 and $60,000 per patient. As the trial advances, the costs rise. A typical phase III trial (where a drug is tested for its value in clinical practice) can cost between $70,000 and $125,000 per patient [15]. As a result, pharmaceutical companies try to avoid small or targeted trials, which are problematic to them for several reasons. First, although a small trial might be more likely to be successful, it implies a small market size and drug companies are sensitive to making large investments if the end market is not big enough to support the up-front cost. Second, the more targeted the trial, the harder to find and enroll patients and validate their results. As a result, some potentially effective therapies might not even reach phase III because it is too slow and costly to conduct the trial.

The perplexing fact, however, is that because of cancer's heterogeneity, a large collection of very specific drugs might be exactly what we need; and the more specific, the smaller the relevant market becomes. It is likely that to maintain therapeutic effectiveness, drugs will have to become increasingly tailored as they target smaller subsets of the patient population. This is a natural consequence of cancer heterogeneity and the long tail problem, whereby sets of increasingly rare driver genes, in aggregate, are crucial in driving the cancer's growth in a substantial number of patients. As a result, if we are ever going to combat these "mutations of unknown significance" in genes that are currently not oncogenes but might one day be recognized as true drivers, our clinical trials will have to become narrower, more specific and molecularly driven in order to maintain their effectiveness. Basket and umbrella trials offer a possible solution to this problem, assuming the baskets can be dynamically subdivided according to the latest molecular sub-classifications emerging from the cancer genomics community.

Combinatorics in treatment pathways

Clinical trials determine which new investigational drugs are allowed to enter medical practice, but the unit on which value is measured in a clinical setting is not a drug, but a treatment pathway. Combination therapies hold even greater promise if we can use data and analytics to prescribe custom, genome-guided cocktails that are tailored to each patient's unique molecular composition. For example, we have already seen progress with the combination of Palbociclib and Letrozole which doubles the progression free survival rate in certain breast cancer patients. Combining Ipilimumab and Nivolumab increases survival by nearly five months for melanoma patients. In multiple myeloma, the combination of Revlimid, Velcade, and Dexamethasone became the standard of care in 2010, after producing such dramatic results that nearly all patients who received the cocktail went into remission [15]. Increasingly, cancer therapies will be administered in combinations and pathways that add a final layer of complexity to the quantification of their effectiveness. In order to accelerate and disrupt cancer care we must integrate the whole process: from the discovery of novel driver genes in primary tumors to tracking the value of treatment pathways in patients. Despite its imposing difficulty, all aspects of the cancer problem hinge on the access and aggregation of molecular, phenotypic and clinical data.

A Data-Driven Solution

Unify siloed data sources

As previously stated, in TCGA and ICGC alone, we have collected about 15,000 exomes to date. Yet this data is devoid of rich therapeutic and phenotypic data collected and observed over long enough patient timelines to be of optimal clinical use. It is akin to collecting 15,000 different locks and trying to unlock them with a handful of keys. We can clearly see, especially in primary tumors, the vast mutation set that is affected at a genomic level when someone develops cancer, yet we cannot follow these patients as they are being treated and see how their tumors respond and adapt to different forms of therapy, allowing us to track the evolution of their disease as they regress or metastasize.

For the first time in our history, we have the ability to generate, store, and analyze genomic data affordably at scale. The cost of sequencing a patient's genome 15 years ago was roughly $100 million; today, that same sequencing can be performed in days for around $1,000, representing a million-fold decrease in the cost of collecting genomic data [16]. This means that we are able to

sequence the 100,000 patients that might be necessary to uncover the majority of primary tumor driver genes for less than the cost of the original human genome. Similar advancements in the underlying cost effectiveness of big data allow us to peer into the vast array of therapeutic data we have collected. Each year, nearly 1.7 million new cases of cancer arise in North America, and roughly 17 million people are treated in totality [5]. With the propagation of EMRs, there is a large volume of patient data that has been digitized and can be mined using NLP techniques and machine learning. By marrying these two large and growing datasets (omics data on the one side collected from sequencing DNA, RNA, protein, etc. and phenotypic and therapeutic data on the other side extracted from EMRs), we should be able to amass a dataset large enough for us to see patterns emerge that were historically invisible to standard analytical techniques. We have the capability to gather unprecedented amounts of data and utilize intense computing to understand cancer at both the macro and micro scale; to prospect the boundaries, peaks and valleys of the cancer landscape.

The primary challenge lies in aggregating the data. First, we are talking about a lot of data. A whole genome can be ~200 Gb of data per patient; so to amass a library of exomes and genomes, we need to inexpensively capture and store genomic data. Second, patient data may be in digital form residing in an EMR system, but the most useful data is generally unstructured and these data sets are growing at a rapid rate. It is estimated that medical information is doubling nearly every five years, and much of the data is in free text fields that are both hard to query and devoid of a common language schema (in other words, doctors do not always use the same words when describing something). That said, the data can be collected, cleaned, stored, and normalized.

Provide deep genome sequencing at scale

As we have discussed, the inherent difficulty of cancer biology requires large volumes of patients and a rich patient characterization, not just at a genomic level but also at a phenotypical level. But, even if we collect a million data points for each patient, if these are too noisy or are missing essential attributes, the size of the data will inhibit rather than enhance our approach. Any data collection effort needs to satisfy the statistical power requirements for identifying the driving genomic factors in a given cancer while ensuring an accurate, rich and reproducible representation of the patient's disease. This begins with the need for a low-cost sequencing solution that is universally available in both the clinical and research setting. We need to collect

large amounts of DNA mutation data (a panel of a few hundred genes is just too small), and of transcriptome data, concurrently. To understand protein dysregulation and interaction, we need to gather data on both DNA and RNA. At the same time, the data we gather has to be of high quality, which in the world of sequencing means high depth of coverage (ideally at least 150-200x for whole exome, and 400-500x for larger gene panels).

Probe unknown drivers beyond gene panels

Gene panels are the current de facto standard in clinical cancer sequencing, as they are effective at revealing mutations in genes that are relatively well understood. However, gene panels limit the opportunity to learn beyond the biology that is already known. There are much richer sources of information such as exome or whole genome sequencing, not to mention the contributions that could be made by measuring the proteome, epigenome, and microbiome. Many more samples are sequenced for genomic DNA than for products of the genome, like RNA transcripts and proteins, because both transcriptomes and proteomes are mutable and noisy, as opposed to DNA whose mutations are permanent and encode the natural history of the tumor; not to mention that adding these other forms of sequencing is more expensive. That said, while panels can be effective tools, especially in the short term for gathering genomic data at a coverage level that is high enough to allow in depth analysis, they are far from sufficient to fully understand the complexity of a tumor. We need to collect broader sets of data, and we need to collect it across a large population of both prospective and retrospective patients. It is imperative that we operationalize and standardize the collection of molecular data, so it can be combined with EMR data for analysis. For this to occur, we need to make sequencing a routine practice in hospitals and cancer centers. It cannot be a “nice to have”, it has to become a “must have”. If we can find a way to sequence a large enough number of cancer patients, then we can build a truly transformative data repository in just a few years.

Characterize patient phenotypes and timelines

This unprecedented volume of molecular data (assuming we can find a way to capture it) will provide the statistical power to map the effectiveness of therapies as a function of patient genomes. Understanding the spatiotemporal dynamics of the tumor genome is especially critical in mapping its response to different therapeutic agents. Here we are not looking for single signals of therapeutic selection across many genes (n to 1) but associations between a myriad of genes and biomarkers with hundreds of therapeutic agents (n to m). Furthermore, both genes

and therapies determine response in a combinatorial fashion, requiring a large number of patients, so as to be able to slice the data over thousands of attributes and combinations. By marrying molecular data with phenotypic and therapeutic data, combining both genomic and clinical markers, a new platform could emerge that supersedes both basket and umbrella trials in its ability to identify response in both large and highly specific patient populations, across groupings of any type.

Discover patterns in patient response

As we amass data, we will begin to see patterns emerge. The addition of rich clinical data to phenotypes and outcomes will allow us to identify previously unknown associations. For example, which mutations in circulating tumor cells in blood lead to metastases in the lung? If metastases occur in the brain, what is the clonal structure in the tumor and does it vary based on the the primary tumor? When a primary prostate tumor is treated with drug A, how many types of resistance exist, and which drugs seem to promote or inhibit them? As a breast cancer tumor metastasis produces new clones, what makes them susceptible to a particular drug and do they vary depending on whether or not the patient was a smoker, or had a thyroid condition, or was taking diabetes medication? As a lung cancer tumor spreads to the bone, does its fundamental circuitry and proteomic profile continue to align to a lung cell, or should we consider reclassifying the patient based on the molecular signature of the new tumor? These are just some of the questions that we can begin to answer once we have collected and analyzed the appropriate variants within our new data set.

Rigorously validate novel insight

As treatment hypotheses are generated, we will inevitably need to test them in a biological setting (synthetic or living) in order to validate our findings; biological systems are far too complex to assume that all discoveries made in a computer dry lab can be replicated in an experimental wet lab, let alone in a patient. In order to properly replicate the complexity of a tumor and its microenvironment, we must attempt to mimic its three dimensional structure so as to account for the interactions between cancer cells and the surrounding normal tissues, or find more effective tools for *in vivo* modeling. One way this can be accomplished is with 3D cell cultures, organoid systems and perhaps eventually with 3D printing of cells. Furthermore, we need to supply the counterforce of an immune system to understand the many evolutionary drivers in the tumor. Injecting cultured tumor cells into mice to create a PDX can replicate some

of the aspects of the tumor's biology within the patient's body. However, key factors are missing, especially since the mice used for growing PDX tumors typically have no immune system. *In vivo* models can be humanized to varying degrees of fidelity, but *in vitro* models, especially microfluidic organs-on-chips and organoids, offer perhaps the greatest hope for testing environments that match a patient at scale. The combination of cost-effective *in vivo* modeling, and advancements in 3D organoids and microfluidics, will provide a host of new tools to validate therapeutics in both the clinical and research setting.

Accelerate learning of cancer biology

In a world where doctors are making decisions with the assistance of massive datasets that are impossible to interpret without the aid of a computer, a system must be put into place that dynamically learns with each new set of data points. This is the realm of Artificial Intelligence (AI). AI is not a single approach; it can use a combination of probabilistic methods, statistical classifiers and deep learning, to extract patterns over structured and unstructured data. Using this smorgasbord of computationally intensive tools, AI has taken an expanding role in almost every field where there are copious amounts of data. For example, the current resurgence of neural networks (more frequently referred to as "deep learning") has been partly due to algorithmic improvements but to a great extent to increases in computational power and expanded learning corpora. Image and speech recognition are two research areas that have benefited spectacularly from deep learning approaches. Face and voice recognition software are now pervasive and perform at levels that are uncannily similar to those of a human.

In contrast to images and speech, machine learning in cancer genomics is still nascent. One of the primary requirements in effectively teaching a machine to learn is knowing the right answer beforehand. With an image, we are teaching a machine to learn the patterns that we recognize effortlessly and once the machine output is produced, we can easily validate that the results are correct. With cancer genomics, we cannot be sure we have the information necessary to even begin to solve the problem. However, there are reasons to believe that deep learning is well suited to cancer genomics. Deep learning works well with signals that are compositional hierarchies [17]. As cells integrate stimuli, information flows and converges until it reaches a handful of molecules that determine the state of the cell. The operations guiding this hierarchical integration, and its compositionality, are still to be resolved, but deep learning has the potential to help us organize the vastness of these biological interactions. If we see even the most modest

degree of learning, perhaps by identifying small sets of extreme responders across dozens of hospitals, our efforts will quickly translate into hundreds of lives being saved.

Conclusion

In our daily lives, we have the power of technology at our fingertips. Algorithms, in the form of cutting-edge analytics and sophisticated software systems, help us navigate our social communities, explore books, music and movies, and search for local businesses and activities right from our phones. However, technology has not permeated healthcare, and in particular cancer care. Many cancer patients are still treated with a one-size-fits-all approach that is eerily similar to the way patients were treated many decades ago.

If we hope to have an impact on the nearly 1.7 million people who will be newly diagnosed with cancer this year in the United States, we need to disrupt the system; and that disruption begins by assembling all the necessary components of an operating system that unifies the collection and analysis of clinically relevant data.

When the personal computer was first built, it was just a pile of sensors and circuit boards, until someone wrote the first operating system. That system connected the keyboard and the screen, it fired up processors when the power was turned on, and provided an interface that allowed the user to program the machine and create their own algorithms. It unified a bunch of disparate functions into a cohesive experience.

We need a similar system in cancer care today. A system that connects anatomic pathology with molecular pathology, and genomic data to therapies and outcomes. A system that communicates bioinformatics and computational biology outputs to a patient's physician or tumor board. A system that integrates validation and modeling with its analytics engine.

This is exactly what Tempus has built: a system that unifies disparate technologies and isolated activities and provides the basic technology infrastructure to offer a seamless experience to a varied array of users, especially physicians that need to consider numerous options to dispense personalized care. In other words, an operating system to battle cancer.

We have recruited a team of accomplished geneticists, data scientists and engineers who have developed software and analytic tools that work within a hospital's existing infrastructure to

augment the care that physicians are able to provide – arming healthcare providers with data and insights to help them make real-time, data-driven decisions.

We have built a platform with the capacity to analyze the molecular and clinical data of millions of patients fighting cancer. With this, we can provide physicians with the insight generated from those who have come before.

The first step to personalizing medicine is to gather the necessary data one would need to customize care in units of one, which requires a common platform combining rich molecular, phenotypic, therapeutic, and outcomes data. Only through the universal adoption of a truly ubiquitous learning system can we hope to lay the foundation for precision medicine in cancer care.

Acknowledgements

We thank Casey Frankenberger for his comments on the manuscript.

References

1 Reuben, Suzanne H., Milliken, Erin L., Paradis, Lisa J., "The Future Of Cancer Research: accelerating scientific innovation, President's Cancer Panel Annual Report 2010-2011", http://deainfo.nci.nih.gov/advisory/pcp/annualReports/pcp10-11rpt/FullReport.pdf

2 Lawrence, Michael S., et al. "Discovery and saturation analysis of cancer genes across 21 tumour types." *Nature* 505.7484 (2014): 495-501.

3 MD Anderson Cancer Center Personalized Cancer Therapy Knowledge Base for Precision Oncology, https://pct.mdanderson.org/#/

4 Vogelstein, Bert, et al. "Cancer genome landscapes." *Science* 339.6127 (2013): 1546-1558.

5 American Cancer Society. Cancer Facts & Figures 2016. Atlanta: American Cancer Society; 2016.

6 NCI's Genomic Data Commons (GDC), https://gdc.cancer.gov/

7 Vice President Joe Biden, “What I Said to the Largest Convening of Cancer Researchers in the Country Yesterday”, https://medium.com/cancer-moonshot/here-s-what-the-vice-president-said-to-the-largest-convening-of-cancer-researchers-in-the-country-3007bb196dbd#.siyj9xn71

8 Sinsky, Christine, et al. “Allocation of Physician Time in Ambulatory Practice: A Time and Motion Study in 4 Specialties” *Ann Intern Med.* (2016) doi:10.7326/M16-0961

9 Muir, Paul, et al. “The real cost of sequencing: scaling computation to keep pace with data generation.” *Genome biology* 17.1 (2016): 1.

10 Hanahan, Douglas, and Robert A. Weinberg. “Hallmarks of cancer: the next generation.” *Cell* 144.5 (2011): 646-674.

11 Gingras I, “The role of precision medicine in “real-life” management of breast cancer patients: A survey assessing the current use and attitudes towards tumor molecular sequencing in clinical practice.” [abstract]. In: *Proceedings of the Thirty-Eighth Annual CTRC-AACR San Antonio Breast Cancer Symposium: 2015* Dec 8-12; San Antonio, TX. Philadelphia (PA): AACR; Cancer Res 2016;76(4 Suppl):Abstract nr P6-04-13.

12 Kiberstis, Paula A., et al. “Metastasis: An evolving story.” *Science* 162 (2016): 163.

13 Scannell, Jack W., et al. “Diagnosing the decline in pharmaceutical R&D efficiency.” *Nature Reviews Drug Discovery* 11.3 (2012): 191-200.

14 CenterWatch FDA Approved Drugs for Oncology, https://www.centerwatch.com/drug-information/fda-approved-drugs/therapeutic-area/12/oncology

15 Scarlett, Uciane K., et al. “High-Throughput Testing of Novel–Novel Combination Therapies for Cancer: An Idea Whose Time Has Come.” *Cancer Discovery* 6.9 (2016): 956-962.

16 The Cost of Sequencing a Human Genome, https://www.genome.gov/sequencingcosts/

17 LeCun, Yann, Bengio, Yoshua, and Hinton, Geoffrey. “Deep learning.” *Nature* 521.7553 (2015): 436-444.

By Eric Lefkofsky, Co-Founder and CEO @TempusLabs and Kevin White, President @TempusLabs

This year roughly 50,000 people will be diagnosed with pancreatic cancer in the United States. With the beginning of Pancreatic Cancer Awareness Month, we have an opportunity to take a look at what needs to be done in order to give the men and women living with pancreatic cancer a fighting chance to survive. According to American Cancer Society statistics, in 2016 pancreatic cancer will claim close to 42,000 lives and it will be the third most lethal cancer (behind lung and colon cancer). With a five year survival rate of less than 8% overall, and a survival rate of less than 2% for people diagnosed with stage 4 disease, a diagnosis of pancreatic cancer gives a patient extraordinarily long odds.

But there is is reason to be optimistic.

In recent years, thanks to the tireless activism and fundraising among pancreatic cancer advocacy groups and research sponsoring foundations, this situation has begun to change. Simultaneously, new technologies such as next-generation genome sequencing have allowed us to catalog the genetic mutations and molecular abnormalities that drive pancreatic cancer. Optimization of protocols for developing patient-derived xenografts and organoids have allowed us to grow a patient's own tumor in a mouse or in engineered synthetic environments, and to study how it behaves when exposed to drugs or to natural conditions it might encounter in its own 'microenvironment.' Meanwhile, advances in immunotherapy and targeted therapeutics have begun to transform the treatment of other forms of cancer, and there are high hopes that at least some of these advances will migrate into the treatment of pancreatic cancer.

In response to this change in climate, our understanding of pancreatic cancer has accelerated at an impressive pace. Last year alone there were nearly 6,000 new scientific references registered in the National Library of Medicine's PubMed catalog; by comparison this is more than double the number published a decade ago in 2005. At the same time, clinical trials have also surged. A search of ClinicalTrials.gov reveals over 700 open studies this month that involve pancreatic cancer, with many of the studies exploring novel therapeutics that harness the immunological or molecular state of the tumor.

The advances happening, driven in part by new technologies, have the potential to be transformative. For example, in 2016 we have learned from genomic analyses that there are at least four molecular subtypes of pancreatic ductal adenocarcinoma (the most common and lethal form of pancreatic cancer) (Bailey et al. March 3, 2016 *Nature*), and that the accepted model of pancreatic cancer development may need to be significantly modified (Notta et al.

October 20, 2016 *Nature*). These findings could have profound effects on how we diagnose, categorize and treat this disease.

The refinement of our molecular understanding of pancreatic cancer has implications for how we might conduct and interpret clinical trials. Recently a new type of clinical trial in pancreatic cancer was announced that matches each patient's molecular profile to their treatment and that can evolve over time based on the data. The Precision Promise trial has been organized by the Pancreatic Cancer Action Network (PanCAN) who is investing $35M across 12 major cancer centers in the United States. Our company Tempus is excited to have been chosen by PanCAN to support this revolutionary effort, using our technology platform for the production and analysis of genomic and clinical data associated with Precision Promise.

For most of the modern world there are algorithms, cutting-edge analytics and sophisticated software systems that help us communicate and navigate in our daily lives, that manage massive financial networks, or that optimize complex industrial workflows. However, this kind of technology has not permeated healthcare, and in particular cancer care, in the same way. We need a new set of technology tools that unifies all of the disparate systems that clinicians and research rely on to collect and analyze data.

To utilize these tools a sufficient number of patients is needed with deep clinical and molecular data, tying therapy to genomic profile and ultimately to outcome. Precision Promise anticipates enrolling thousands of patients in the next few years. If the technology we have at our fingertips holds the key to advancing our understanding of treatment outcomes and extending survival, Precision Promise (and other trials like it) may be the beginning of the end for pancreatic cancer.

Do we have the technology to cure pancreatic cancer, or to at least improve outcomes so that patients with this dire diagnosis can survive longer than they do now? The answer is maybe, which means there is hope. Many think we are on the precipice of a new age in the treatment of all cancers, including pancreatic cancer. While time will tell, time is unfortunately not something pancreatic cancer patients have to spare. That is why we are teaming up with other like-minded organizations and individuals to see how fast we can learn from the technologies we have in hand today. We need more innovative initiatives like Precision Promise across more cancer types. And if our current technology is going to put a dent in pancreatic cancer, let's figure that out fast. The month of the November is a time to remind ourselves of the urgent need for progress toward conquering this deadly disease.

A UNIFYING PLATFORM

This past year, President Obama and Vice President Biden brought new energy to our country's efforts to end cancer as we know it with the Cancer Moonshot initiative. The goal of Moonshot is admirable – accelerating the progress of cancer research – but even its most fervent supporters know that there are barriers to achieving success.

First, collaboration between institutions, researchers and industry has long been an obstacle to moving the needle faster, in part because the healthcare system is set up to reward those who make diagnostic and drug discoveries. We reward patenting and publishing. With limited public funding and philanthropic support, investigators are often motivated to build walls around their research and silo their data instead of sharing ideas and democratizing their findings.

Second, highly regulated industries tend to adopt technology more slowly than non-regulated ones. As a result, cancer research has not kept pace in integrating information technology and modern software analytics that have impacted so many other sectors of our economy. The core technology platforms that are prevalent in most major hospitals are antiquated by today's standards, leaving physicians without the tools they need to quickly access and analyze critical information.

Which brings me to the third barrier: data. There is simply not enough data for researchers to analyze and for clinicians to work with to affect change. While personalized medicine is happening in isolation, it is nearly impossible to scale these efforts without vast amounts of phenotypic, therapeutic, and molecular data. As of today, the two largest combined public data sets include data on roughly 20,000 patients, a tiny portion of the nearly 50 million people who are living with cancer worldwide. And when you break these data sets down by cancer subtype, you are often left with a few hundred patients in totality; far too few to produce statistically significant patterns.

While the Cancer Moonshot is admirable, there is a structural problem that has to be addressed concurrently if we hope to tackle a disease that has the fortitude of over 50 million years of evolution on its side. The healthcare system needs an independent data and analytics platform that physicians can connect to and utilize in both a research and clinical setting. In other words, we need an Operating System for cancer.

The first step in building such a system is to foster the collection of molecular data. We need to take genomic sequencing and molecular profiling from being a tool largely used for research to one that becomes the status quo when a patient is diagnosed with cancer (especially late-stage or metastatic cancer). The second step is to combine molecular data with vast amounts of phenotypic, therapeutic, and outcome data extracted from electronic medical records, so we can analyze patients at a far more granular level looking for clinically relevant patterns. And the final step (for now) is to build a scalable technology platform that allows physicians to test different therapies digitally, as well as biologically through in vivo and in vitro modeling, to see which ones might have the greatest impact on a patient's disease.

I've built my career by using software, data and analytics to disrupt industries such as printing, logistics, media, manufacturing and local commerce. If we hope to have an impact on the nearly 1.7 million people who will be newly diagnosed with cancer this year in the United States alone, we need to disrupt the system; and that disruption begins by assembling all the necessary components of an Operating System that unifies the collection and analysis of clinically relevant data.

That's the goal behind Tempus, a company I launched about a year ago to enable physicians to deliver more personalized cancer care by using the best attributes of big data and machine learning as tools they can harness when treating their patients.

We have recruited a team of accomplished geneticists, computational biologists, data scientists and engineers who have developed software and analytic tools that work within a hospital's existing infrastructure to augment the care that physicians are able to provide – arming oncologists, pathologists, radiologists, and surgeons with data and insights that can help them make real-time, data-driven decisions.

We have built a platform with the capacity to analyze the molecular and clinical data of millions of patients fighting cancer. With this, we can provide physicians (and their patients) with the insight generated from those who have come before.

In order to win the battle against cancer we need to bring together technologists, scientists, and physicians to work toward a common solution; all contributing data and insight to a collective system. Only through the universal adoption of a truly ubiquitous learning system (an Operating System for cancer) can we hope to lay the foundation for precision medicine in cancer care.

When the personal computer was first built, in a garage in California, it was just a pile of sensors and circuit boards, until someone wrote the first operating system. That system connected the

keyboard to the screen. It fired up processors when the power was turned on. It ran applications and allowed people to enter in commands. It unified a bunch of disparate functions into a cohesive experience.

We need the same thing in cancer today. We need one system that unifies disparate activities and provides the basic technology infrastructure that physicians need to dispense care.

It's time that we put technology and big data to use in personalizing cancer treatment – giving physicians, patients and their families the fighting chance they deserve.

IT'S ABOUT TIME…

Roughly 15 years ago, scientists first mapped the human genome, leading many to believe that this would usher in a new age of personalized therapies to treat cancer and many other diseases. However, despite tremendous advances in technology and our understanding of cancer, the disease still kills 600,000 people a year in the U.S., accounting for more than one in five deaths.

Even as genomic sequencing has become less expensive and more widely available, cancer patients are treated in much the same way today as they were 25 years ago. The U.S. health care system has not yet found an efficient and systematic way to gather large amounts of molecular data, combine it with an individual patient's therapeutic data and then analyze it in a clinical setting to drive individual treatment decisions.

Why? Because big data does not yet exist in the cancer arena.

The two largest combined public data sets include data on less than 20,000 patients, a tiny portion of the millions of people who are living with cancer worldwide.

Hospitals – even broad hospital networks and cancer centers – don't have the resources needed to sequence enough patients in a cost-effective manner to aggregate molecular data, nor do

they have the capital or technology resources needed to invest in analytics. In addition, even the largest hospitals see a relatively small number of patients, so generating the volume of data needed to find clinically relevant patterns would take far too long.

That's why I started Tempus, a health-tech company that brings technology and big data together to improve cancer care and give physicians the tools they need to personalize treatments for their patients.
We have recruited a world-class team of accomplished geneticists, computational biologists, data scientists and software engineers who have developed software and analytic tools that work within a hospital's existing infrastructure to analyze data and provide decision support for health care professionals whose patients are not responding to conventional therapies.

The traction Tempus has gained in less than one year of operation is an indication that major hospital systems recognize the need for the types of services that Tempus is offering. Today, we announced a partnership with the Robert H. Lurie Comprehensive Cancer Center of Northwestern University and there's more to come.

Our goal is for each patient to benefit from the treatment of others who came before by providing physicians with a technology platform that learns as we forge more partnerships and gather more data.

It's about time we give healthcare providers the tools they need to personalize cancer treatment.

LEFKOFSKY STARTUP TAKES ON CANCER

Published in Crain's Chicago Business, September 29, 2016 –
On a shelf behind Eric Lefkofsky's desk are bound copies of the closing documents of the three companies he's taken public, Groupon, InnerWorkings and Echo Global Logistics. Perched above them is a plastic structure made by a 3-D printer and given to him as a gift, a tiny double helix.

There you have his past and his present.

After making a fortune from his e-commerce startups—Forbes pegs his current wealth at $1.79 billion—Lefkofsky has moved on to his biggest challenge yet. Through his latest venture, Tempus, which he has just taken out of stealth mode, the serial entrepreneur is taking on cancer. His plan: to compile a massive genomic database about the disease that can be compared with an individual patient's own DNA to help doctors personalize therapies. If it succeeds, people could live with cancer longer. Some might even be cured.

By now, of course, Lefkofsky is comfortable with risk. But why, with no background in genetics or medicine, is he self-funding a genomics company?

His wife, Liz, was treated for breast cancer a couple of years ago. She is doing well, taking it one day at a time, he says. He couldn't say the same for the state of cancer treatment. To help her find the best course of action, Lefkofsky learned everything he could about breast cancer and how doctors were treating the disease. He found hospitals as far behind in the use of data as Groupon discovered small merchants were.

"You realize if you've been a patient or know someone who's been a patient that technology has not permeated health care and certainly oncology the way it has permeated other industries," he says.

"At some of the top cancer centers in the country, you find really interesting research initiatives where they're sequencing patients, looking for patterns," he says. "It's not happening at scale."

At the same time, he says, doctors can be overwhelmed. "They're caught in this paradox: You can collect lots of data, but you can't necessarily analyze all the data."

He's not the only one to have this realization, however. Hundreds of startups, as well as IBM through its Watson Group and nonprofits like the Venter Institute, are using Big Data to determine which drugs will be most efficacious in treating complex diseases. And their numbers are multiplying as genetic screening and data storage become cheaper and the amount of information grows.

"I'm encouraged he's doing this. The world needs more of this work," says Immanuel Thangaraj, former chief financial officer of Chicago's Arch Venture Partners and now a managing director at Park Lane Ventures, a Silicon Valley fund that specializes in health care. "But this is a crowded space. He's joining a race like the Boston Marathon."

Lefkofsky says Tempus is unique because it is an end-to-end solution. Through partnerships with leading hospitals, he hopes to collect a massive set of data. In addition to providing gene-

sequencing tests to patients from a 20,000-square-foot lab at its River North headquarters, Tempus gives doctors software that allows them to quickly compare a patient's genetic profile to its database of other cancer patients to determine which treatments have been the most effective, as well as to find clinical trials. Hospitals, in return, contribute data.

Tempus announced its first partnership with Northwestern's Lurie Comprehensive Cancer Center on Sept. 29. It's expected to announce others later this fall.

Lefkofsky, 47, began constructing Tempus late last year after stepping down as CEO of Groupon. (He's still chairman and its biggest shareholder.) He soon hired Kevin White, a geneticist and top researcher at University of Chicago who had spent a decade building its Institute for Genomics and Systems Biology.

Tempus is now up to almost 100 employees, heavy on Ph.D.s. It's a big departure from legions of millennials working the phones at Groupon and other companies.

But Lefkofsky says they have much in common. "If you look at all the businesses we've built . . . everything we work on has a Big Data component." Says Michael Sacks, CEO of GCM Grosvenor, a $45 billion asset management fund, and an investor: "Tempus is similar. It is just much bigger and much more important."

Like all their ventures, Tempus is owned 50-50 by Lefkofsky and Brad Keywell, his longtime business partner. Keywell, too, is doubling down on Big Data through his fast-growing startup Uptake, which is pulling in data from sensors on machines to better forecast service needs for Caterpillar and potentially other industrial clients.

"We don't start companies thinking about the industry or big problems, or hey, let's take a big swing because we can," Lefkofsky says. "When you have five successful companies, it's not like, 'I just really want to start a sixth.'

"Our motivation for starting companies is very personal: We come across some problem, and you have this lightbulb that goes off that says, 'I have this solution.' "

Lefkofsky and Keywell won't say how much they've invested in their latest startups. But the three IPOs and the $750 million sale of another of their startups, ad-billing company Mediaocean, allow them to endow their venture fund, Lightbank, with $200 million. Lefkofsky says he's willing to invest up to $100 million in Tempus.

In its partnership with Northwestern, patients will pay Tempus for gene-sequencing tests and Northwestern doctors will get free use of its software. Northwestern, in return, will provide anonymized patient information to help Tempus build up its database.

“I’m excited to see where this goes,” says Michael Liang, a Chicago-based partner at Baird Capital venture fund. “The biggest question is reimbursement. How do you get paid for these tests and differentiate your tests?”

Lefkofsky says he’s up for the challenge. “I’m confident that in five or 10 years, the average oncologist will be connected to a system like Tempus. Whether or not it’s Tempus, it’s too soon to say.”