Thank you for that kind introduction, and thanks to all of you for giving up your late Saturday afternoon. I’m going to present part of a talk that the Howard Hughes Medical Institute asked me to do. It’s called “The Holiday Lectures on Science” (http://www.hhmi.org/biointeractive/medicine-genomic-era). HHMI spends a lot of effort trying to disseminate science into the high schools in the Washington, D.C. area. The headquarters of the Howard Hughes Medical Institute is in Chevy Chase, so I want to give them a shout-out for this.

The term “precision medicine” was first coined in a report that we published in 2011. A lot of these reports sit on the shelf and gather dust and this one was doing that until January of 2015, when President Obama declared the Precision Medicine Initiative at the White House. It took a little time to get the wheels of government to buy into this. Now, the “Cancer Moonshot” thing is a more recent development. Many of you, I assume, watched at least part of Obama’s State of the Union when, in the middle of it, with no prior announcement, he turned around and looked back at Vice President Biden and said, “It’s time to cure cancer, and I’m putting you in charge of Mission Control.” So, for the last 4 or 5 months, the vice president has been busy living up to that job, trying to get advice on what to do. I’m one of 36 people on a panel where we are on conference calls frequently.

I want to leave you with three main take-home messages today. The first one is that this is an incredibly exciting time for cancer research. The second is that this has been a long time in coming. It required years of commitment to understanding basic cell biology and cancer biology, largely due to work at places like this. The third thing is that the “moonshot” is not one project. There is not a single moonshot. This is a problem we’re going to have to live with for a long time coming, and we have to avoid the hype of most media articles on a new discovery saying, “Oh, aren’t we almost there?” I don’t want to take the wind out of the sails—it’s an incredibly exciting time—but we have to manage expectations.

Let’s start with first principles. You all know the terms “benign” and “malignant.” “Benign” means a tumor that starts in one cell and proliferates inappropriately, but it just sits there; it’s noninvasive. It’s not causing a problem unless it gets really big, and then you just ask the surgeon to take it out. It starts to become a problem when the initial tumor cells acquire the ability to escape the “mother ship” and start invading into other tissues. The real problem is if it gets into the blood, then it can metastasize to other sites. Benign tumors are not cancer. The locally invasive and metastatic tumors are malignant. These are cancer.

How do we treat these? We rely on surgeons to cure cancer by physically removing the tumor cells. Obviously, that is the best way. We have to detect tumors early enough to have that work. The two other main therapies you’ve known about—and you’ve had friends who’ve had these therapies—are radiation and chemotherapy. These kill cancer cells by damaging the DNA in the cells to the point where the cells can’t survive. But because normal cells and tumor cells both require DNA to replicate, these two therapies kill normal cells, too. We need more therapies because a surgeon can’t see cells that have escaped the tumor, and neither can a radiologist. Therefore, surgeons can leave cells behind. Radiation and chemotherapies don’t discriminate well enough between normal and tumor cells; therefore, there are a lot of side effects.

The excitement over the last 15-plus years is from new treatments that have emerged from our knowledge of the DNA sequence of tumor cells. We now know that cancers are caused by mutations in our normal genes. We call these cancer-causing genes, and we’ve now adopted the terminology of “drivers.” These are mutations in a gene that now become the driving force of the cancer. There are two flavors. There are those that we call “oncogenes,” which are “dominant,” in the sense of “Mendel-and-the-peas” dominant versus recessive traits. The other category is composed of recessive “tumor-suppressor genes.”

Cells, when they divide, go through something called the “cell cycle.” An oncogene is a “Go” signal. A tumor-
suppressor gene is a “Stop” signal. If you get a mutation in one copy of a gene that’s an oncogene, it sends a “Go” signal and the cell divides, and you form a tumor. For a tumor-suppressor gene, one mutation doesn’t do it. You have to mutate both copies of the gene. Some of you might know that tumor-suppressor genes sometimes are inherited. The famous examples are the breast cancer genes, BRCA1 and BRCA2. Families who have that mutation in their germline (meaning it’s in every cell in their body and can pass on to their children) have one mutated gene already. Those people are at higher risk to get tumors because they only have to develop one additional mutation; the other one is already gone.

A very simple concept is, if you have a mutant oncogene and you know it’s driving the cancer, why not make a drug that inhibits the mutant protein? That is what’s called “targeted therapy.” Perhaps the best example of this is a drug called Gleevec.

Leukemia is a cancer of the blood that results in the overproduction of white blood cells. The earliest symptoms are anemia, but mostly it’s diagnosed by examining blood and bone marrow smears under a microscope. There are too many white blood cells in patients with chronic myeloid leukemia. You can grow these cells in the laboratory and look at the chromosomes, and if you use a technology called cytogenetics you can line up the pairs of somatic chromosomes—1 to 22—and then identify the sex chromosomes, XX or XY. If you’re really good at looking at these chromosome portraits—these karyotypes—all the time, you can see that one of the pairs of chromosome 22 is too short. This discovery was made in 1960 in Philadelphia, and this is called “the Philadelphian chromosome.” It’s a pretty subtle change, but such is the power of observation.

It took another 10 to 12 years of developing technology to recognize that what was actually happening is that the missing little piece from chromosome 22 is actually now over on chromosome 9. What basically happened in one cell in the blood is that there was a break at the bottom of chromosome 9, there was a break at the bottom of chromosome 22, and these two pieces switched places. This is called “reciprocal translocation.” It turns out that break is on chromosome 22, and these two pieces switched places. This is chromosome 9, there was a break at the bottom of chromosome. What basically happened in one missing little piece from chromosome 22 is actually now on chromosome 9, which goes by the name BCR (breakpoint cluster region), and the one on 9, which goes by the name ABL (Abelson murine leukemia viral oncogene homolog), forms this hybrid gene called BCR–ABL, which creates a cancer-causing protein that is half-BCR and half-ABL. This hybrid gene does not exist in normal cells.

We have solved the crystal structure of this protein, so we actually know what the protein looks like at atomic scale. This protein is a molecular machine called an enzyme, which executes a function. If you look at the “skeleton” of the protein, there’s this loop that can move back and forth between what’s known as an active and inactive conformation, essentially a “flap” that opens and closes a pocket or recess in the protein structure where another molecule might bind. In this case, it binds a molecule called ATP (adenosine triphosphate). It then transfers one of ATP’s phosphates—it has three—to another protein called a “substrate.” That’s a molecular signal. Think of this as a chain reaction, with a bucket brigade passing a signal from one protein to another, with the ultimate result of this chain of signal propagation being cell proliferation. That’s what BCR–ABL does.

Now that we know exactly how this works, we can make a drug that would interfere with its activity? And can we do it very precisely so that it does not interfere with every protein that uses ATP? It turns out that using ATP is a process commonly used by many different proteins—over 500. In Basel, Switzerland, in the early 1990s, some chemists decided to look for such an inhibitor. When they started this project, we didn’t know much of what I’ve just told you. That knowledge came later. You could imagine though, if you knew what I just told you, that maybe you could just draw on the board exactly what chemical would fit in that pocket where the ATP was binding. It turns out that even with what we know now, that’s still incredibly challenging. The approach that was used more than 20 years ago is still the most efficient way to at least get a first sense of what the chemical will look like.

The successful approach came through a brute force strategy that, at the time, was only accessible to the pharmaceutical industry. Imagine searching through a whole library of chemical structures that chemists have made over the years, millions of them, to see if one might block BCR–ABL. That’s exactly what happened, and that’s how Gleevec popped out. Gleevec slips right into this position where the ATP would have been. What’s really clever about it is that it slips in and locks this activation loop in what’s known as the “closed” conformation. Even if the loop tries to move, it can’t; Gleevec completely disables the signaling.

What happened when this drug went into the patients? I think you know a little bit of that story. Six patients who got the drug had white blood cell counts in the $5 \times 10^3$ cells/µL range before they started taking the drug (normally, it’s in the $5 \times 10^3$ cells/µL range). All the blood counts came down to normal or below ranges within 30 days of treatment with Gleevec. Some patients with this disease who were first treated on clinical trials in the 1999–2000 timeframe are still alive and well, taking this pill once a day and living a normal life. That is what “targeted therapy” is. That is what “precision medicine” is: This pill works in patients who have the BCR–ABL mutation, but it doesn’t work in patients who don’t have the BCR–ABL mutation.

This one experiment led to a series of conversations among the whole field. In fact, I remember being in this auditorium about 15 years ago saying, “Is it time to do something really big?” Is there sort of an engineering approach to push this? How many more examples of this are there? It started in 1960 when Peter Nowell was looking through a microscope and saw the Philadelphia chromosome pattern. It was 40 years later when Gleevec was approved. We can go faster. The thing that enabled us to think about going at large scale was the ability to sequence DNA at great speed and low cost, as well as...
the computing power to maintain and process the massive amounts of data related to DNA and protein sequencing. Basically, the idea was to expand the example of Gleevec to other targeted therapies by defining all the cancer driver genes and their mutations.

New technology called “next-generation sequencing” has drastically increased the pace of cancer research. Using the old technology, it cost one billion dollars to sequence the first human genome. Using “NextGen” technology, that cost is down to 1000 dollars. The big idea is that, instead of relying on cytogeneticists to look in a crude way at chromosomes, let’s sequence the DNA base by base—all three billion base pairs in the tumor—and compare it with the patient’s normal DNA to see if we can find mutations specific to tumor cells. Let’s do this for thousands of tumors and see what we learn.

I’ll also remind you that the reason these mutations create a problem is because the DNA sequence, the so-called triplet codon, instructs the cell’s machinery to insert a specific amino acid (for example, the triplet TAC encodes tyrosine) into a protein as it’s being made. A mutation is a misspelling: The “spell-check” function didn’t work. If the third base changes to a T, that’s called a silent mutation because that TAT triplet sequence still codes for a tyrosine. Those silent mutations are screened out from this giant sequencing project to avoid confusion; those mutations are less likely to matter. We’re more interested in the two other types of mutations. In missense mutations, changing the A to a G (TGC), for example, changes the tyrosine to a different amino acid, cysteine. That can have a very profound effect on a protein. The other type of mutation is if there’s a codon change where the C turns to a G (i.e., TAG), it becomes a stop codon; the translation of protein from the messenger RNA that this derives from just completely stops. The protein doesn’t get made properly. That’s a very profound disruptive change.

Through a national effort run by the National Cancer Institute, as well as an international effort, the DNA-sequencing community and cancer biologists got together and sequenced well over 10,000 tumors from 23 different types of cancer. What did we learn? We can generate graphs where we can count the number of mutations we found in a tumor and we can now start to look at some patterns. There’s something really interesting that just grabs you right off the bat, and that is that children with cancer have tumors with very few mutations: only four or five per tumor. Adults have more and there’s a wide range, from 25 to hundreds to thousands, although most tumors have 50 to 100 mutations.

What’s underlying this pattern? We think of lung cancer as lung cancer, and we know that smoking is associated with lung cancer. If you get lung cancer and you weren’t a smoker, you have very few mutations. If you have lung cancer and you are a smoker, you have 150 to 200 mutations. This makes perfect sense. We know that carcinogens in tobacco smoke cause DNA mutations. We’re not learning anything that we didn’t suspect, but now we have real quantification of this. Patients with melanoma also have tumors with 100 to 200 mutations. Again, once we saw this, it made perfect sense. Skin cancer is caused by exposure to sunlight. Ultraviolet light causes DNA mutations. That’s why we wear sunscreen. The tumors with the mutations in the thousands, it took a little while to sort this out, but this is primarily a subset of patients with colon cancer who have an underlying mutation in genes that are part of the “spell-check” (i.e., the cellular machinery responsible for maintaining the integrity and fidelity of the gene sequence). If you disrupt the “spell-check,” the error rate is going to go way up, and that’s exactly what happens here, and it turns out this makes a big difference for therapy as well.

What I’ve just described to you is the number of mutations, but how do we know if those mutations are in genes that are important in causing cancer? Are they drivers? How do we sort that out? We still don’t know the best way to do this, but a very good way is to count them, determine where on the chromosome they are, and in what genes they are. The mutations can be laid out on a three-dimensional plot that displays each chromosome from 1 to 22 and then the sex chromosome from left to right (i.e., the x-axis). The length of each chromosome is displayed along the y-axis, with the start of each chromosome at the origin; the prevalence of mutations in a given gene on that chromosome is displayed vertically on the z-axis. You see that there are these skyscrapers of mutations, these towers. These are the genes that we know—and in many cases, already knew—are cancer genes. The stacks that are shorter in size are also clearly cancer genes because we keep seeing them again and again, but they are less common. In this way, as a statistical strategy, we can determine which genes are most likely to be the true drivers.

Now that we’ve analyzed so many thousands of tumors, where do we stand? It turns out there’s quite a few driver genes. The most commonly mutated gene is called TP53 (tumor protein 53). Some of you might have heard of other commonly occurring mutants: BRAF, KRAS, PTEN, PIK3CA. We can’t really know ever if we have identified them all, but we certainly know we have the common ones, where “common” is defined statistically as present in at least 5% of the tumors of that type. If we sequenced another hundred thousand tumors, there’s no question that we would find more, but these new ones will be less common across the population of cancer patients. The good news about this is that we know our enemy in way better detail than we did 10 years ago. It’s challenging, but it’s only several hundred genes, not thousands. As you’ll see, we can bucket these hundreds into smaller groups to which, hopefully, common therapies can be applied. There’s no doubt that cancer diagnosis will include—and is already including—sequencing to find out what’s the status of these genes in a tumor, because it really matters.

One way of bucketing those hundreds of genes into just a few categories is to group them based on their function in normal cells. The smallest group consists of the genes involved in genome maintenance, which is the “spell-check” function in the cell. Of the other two major groups, one is simple to understand: It’s the “Go” signal
that I mentioned before for cell growth and survival. EGFR (epidermal growth factor receptor) is an example of a growth and survival gene. It’s mutated in several different cancers, primarily lung cancer, and there are drugs that now have changed the treatment of lung cancer based on that. EGFR starts that ATP signaling process: When its ligand, epidermal growth factor, binds to the receptor (which is embedded in the cell membrane), it initiates a chain reaction of ATP phosphorylation of all these downstream proteins (RAS, RAF, MEK), causing cells to grow. Other growth-related cancer genes are drivers on the oncogene side like cyclin D1 and CDK4 that stimulate the cell cycle, and on the tumor-suppressor side, P53 and RB, which inhibit the cell cycle. We understand the function of these in exquisite detail.

The other major category is genes that are normally involved in what’s called “cell fate.” All tissues—blood, pancreas, liver, prostate—begin with a single cell called a stem cell. These produce progenitor cells, which then differentiate. Blocking any of these steps can cause cells at a particular stage of development to overproliferate and form a tumor, possibly a malignant one. The idea is that there’s a set of genes that control this process in a very regulated way. What we’ve learned from these sequencing projects is that, if one of the genes that control this goes awry and the cell can’t differentiate, you’ve got a problem.

One of the best examples is colon cancer. The intestine is lined with a series of finger-like projections called villi, and around the base of each villus are “crypts,” deep recesses in the lining of the intestine. The stem cells at the base of each crypt give rise to their “children”—the progenitor cells—that move upward out of the crypt, where they undergo their final differentiation to the cells covering the villi. These cells are critical; their function is primarily to absorb food nutrients. These cells actually only live for a couple of days; there’s massive turnover. When they get to the top of the villus, they’ve hit the end of their useful life and they die in a process called programmed cell death. What happens if you have a mutation in one of these cell fate genes that’s important in this process? There’s a famous gene called APC (adenomatous polyposis coli) that does just that. If it’s mutated, cells that are supposed to keep progressing out of the crypt never do. Because the APC gene is lost, they can’t complete the migration and differentiation process and they pile up in the crypts, and you get colon cancer.

How many of these hundreds of mutations can we deal with? As an example, lung cancer has been a remarkable success story. For decades, lung cancer has been diagnosed through the microscope. We still do that. A surgeon takes a biopsy; we then look under the microscope and we see these different patterns, and those have been the basis for diagnosis and classification. Nearly half are defined as adenocarcinomas, followed by squamous and small-cell cancers, and finally, large-cell cancer. With DNA sequencing, we now know that just one histological pattern, adenocarcinoma, has numerous different types of mutant genes, so lung cancer is actually several diseases with different cancer-causing genes. There’s still a por-

tion we haven’t yet figured out, but we know more than half of them. KRAS and EGFR each are mutated in about a third of adenocarcinoma tumors, with another third composed of a host of less common mutations (e.g., ALK fusions, HER2, BRAF, PIK3CA, AKT1, etc.). Clearly, “adenocarcinoma” is way too vague and broad a category; we need to subdivide these into the drivers responsible for them. The good news is that there are drugs either already available or that have been discovered over the last several years that inhibit many of these because they share that property of BCR–ABL of being an enzyme that requires ATP. In the same way that Gleevec was discovered, the inhibitors of many of these mutant genes have now also been discovered.

Does it really work? Here’s an example of one of the first patients who received one of these pills that inhibits EGFR. This is a chest X-ray. All of this fluffy material is tumor and, within just a matter of a week, this patient’s tumor was clearing away. This is a remarkable story. Now when we do clinical trials, we look at the results in graphs called waterfall plots. Each bar along the x-axis is a patient, and what’s being measured on the y-axis is how big was the tumor before they started the therapy and how big was it a month or two later (essentially, the increase or decrease in tumor size); the patients are arranged from left to right along the x-axis from worst to best responders. With the EGFR inhibitor drug, erlotinib, there are patients who’ve now had almost 100% shrinkage of their tumor (and a lot of them), but there are others that haven’t. A big part of what we’re doing in the field is trying to understand, if all of these patients have the mutation, how come everyone didn’t benefit? We’re getting lots of new insights leading to the combinations of drugs that move this waterfall plot in a favorable direction (i.e., toward the right).

What I want to end with is this new form of therapy—I say “new,” even though it was first tried a hundred years ago—called immunotherapy. To introduce this, I want to remind you how the immune system works. Imagine that you get a viral infection. It gets into the cell. It starts producing viral proteins in the cell. When a protein is made, it has a certain length of time that it’s useful and then it degrades. When the cell degrades it, it turns out that these little peptide fragments of the viral protein then get presented by a molecule on the surface of the cell called a major histocompatibility complex (MHC) protein. There’s a whole code of the shape of the peptide and the shape of the MHC that’s very important in controlling the precision of the immune response.

In a similar way, you could imagine that a cancer cell that has a mutant driver, it’s not normal. Could the immune system recognize that protein? Does it also get chewed up? Does it get presented on the outside of the cell? We now know that all of these are true. A T cell, which is part of the immune system, has a receptor that very specifically recognizes and binds to the viral peptide in the MHC on the surface of an infected cell. When that happens, that binding triggers the proliferation of the T cell so that it will kill the infected cell. That’s how we get rid of an infection. If cancer were an infection, the same
thing should happen: The immune cells should proliferate and kill the cancer cell before it spreads. We know that’s not the case, so what happened? It turns out that, even in a normal situation, there’s a built-in sort of fail-safe or backup system to control the immune response. If we didn’t have a way to turn it off, the immune system, when overactive, can be lethal. There are a series of molecules on the surface of the cells called “immune checkpoint molecules” that turn the immune system off once it’s been activated. There is exquisite detail in how this interaction works. We now know that the reason these cancers aren’t eliminated quickly is that cancer cells have figured out a way to borrow that strategy and turn off the immune system before it even gets activated.

What can we do about that? Can we figure out a way to block this interaction and outsmart the tumor cell? There are a number of drugs that have recently been approved that go by the name “checkpoint inhibitor.” One is an antibody against a checkpoint molecule called CTLA4, and another is against another checkpoint molecule called PD1. These two drugs—and there are many versions; different companies have developed them—are the most widely studied cancer drugs today. These drugs are being studied in every tumor type, and they are now approved as treatment for at least three or four different types. When you give the inhibitor, you disable the checkpoint molecule, and the T cells proliferate and the tumor cell is killed. The first example of that working was in melanoma. Remember, melanoma is caused by sun exposure. Those are tumors that happen to have a lot of mutations. This melanoma patient with metastases to the liver received CTLA4, and 36 weeks after the start of treatment, the melanoma disappeared, even in the liver. One of the most famous examples of this treatment being successful is former President Jimmy Carter. He had melanoma with metastasis to the brain. He got one of these drugs, and his disease completely cleared. What’s really amazing about this is that some of these patients are out more than 10 years without a relapse. The immune system in this case has been “taught” to recognize cancer cells, and that “knowledge” is still there. It’s like a drug that’s continually active to pick up any further tumor cells.

Now, remember, the patients with lung cancers that didn’t smoke have a low number of mutations, but the ones that did smoke have a high number of mutations. It turns out that these immune drugs work best in melanoma and in lung cancer, but only in the patients who have a smoking history. The other cancer types with very high numbers of mutations—the “spell-check” tumors, the colon cancers—also respond to this category of drugs. This is late-breaking in terms of why this is happening, but the leading hypothesis—and I think it’s a very compelling one—is that, if you’ve got a lot of mutations, the probability that you’ve created a mutation in a peptide that’s going to stimulate the immune system is higher. Again, having a lot of mutations is clearly bad, but your probability of responding to one of these drugs is increased if you have a lot of mutations.

So, now we’ve got surgery, radiation, chemotherapy, targeted therapy, and immunotherapy. We have an incredibly exciting arsenal. This is why all of us spent 5 days here talking about the nitty-gritty of how these drugs are working. The pace at which these drugs have been developed is startling. Gleevec was approved in 2001. It wasn’t the first targeted therapy, but at the time, you could count them on one hand. It was pretty impressive just 2 years ago, when we were up to 35, but I just looked online yesterday, and there are 71 approved targeted therapies. It’s a testament to this massive investment in biology, sequencing, etc. It’s also a testament to the power of the biotechnology and pharmaceutical industry to react to this investment in public science and take it to the next level.

What’s left to do? First, I think I’ve made it abundantly clear that this has been possible only because we understand the problem in the detail that we didn’t have back when the “War on Cancer” was declared by President Nixon. We understand the taxonomy of how to classify tumors. We’re using mutations to determine what cancer the patients have. We understand a lot more about how resistance is developed, etc. All this has led to an increasing number of drug and other treatment options, but we still have a lot of work ahead, and I want to finish with just a few thoughts on what that work looks like.

Together with my colleagues at Sloan Kettering, led by David Solit and Marc Ladanyi, about 2 years ago, we started sequencing the genomes of our patients who have metastatic cancer. We’ve done a little over 10,000. In two-thirds of those patients, we know what caused their cancer, but in a third, we still don’t. We see a few mutations, but they’re not on the driver list. We call that “cancer of unknown driver,” and we need to figure that out. Otherwise, we’re not going to have a good strategy to treat those tumors. We can talk about how we think we can do it, but it’s not trivial. It’s a challenge. The other problem is that, even of the two-thirds where we know exactly what’s going on, we only have drugs for about half of them. The other half, which we call “undruggable,” has not been neglected because these are rare cancers that do not attract the interest of the pharmaceutical community. These are super-challenging. These drivers are not these ATP-dependent enzymes. We actually don’t know how to go after them chemically. Of course, we didn’t know how to go after these ATP-dependent enzymes chemically until Gleevec. I’m optimistic. We just have to keep investing.

I’ll leave you with the idea that there’s not a single moonshot project that finishes the job. It’s still complicated. I’m very grateful to the vice president for bringing so much attention to the problem. He has galvanized conversations among many branches of government that weren’t happening before and among the pharmaceutical sector and among scientists. I’m an eternal optimist. We are going to continue to make progress in this fight.

Thank you very much.