Elizabeth McKenna: Can you tell us about the efforts to identify mutational signatures from cancer genome sequences?

Sir Michael: We’ve known for about half a century that all cancers are caused by somatic mutations: mutations that take place in all cells of our body throughout the course of our lifetime. Remarkably, we know relatively little about the causes of those mutations. We know, for example, that smoking cigarettes damages the DNA of the lung and that damage gets converted into mutations, and those mutations cause lung cancer. Similarly, we know that exposure to ultraviolet light from the sun causes damage in cells of the skin, and that damage gets converted into mutations and can result in skin cancer. However, we have only a rudimentary idea as to what causes the mutations in breast cancer, or glioma, or pancreatic cancer.

One possible way to elucidate this problem is to actually look at the cancer genome itself. A series of studies in the 1980s and 1990s looked at one particular gene, the \( p53 \) gene, across a range of human cancers and found that the pattern of mutations gave insight into the cause of the cancer. If one looked at the \( p53 \) gene in lung cancers, one found a particular type of mutation, a C going to an A being predominant. Conversely, if you looked at melanomas or squamous carcinomas of the skin, you saw C to T mutations predominantly. This made a certain amount of sense, because it was already known that ultraviolet light causes the sort of damage that would result in C to T mutations, and that mutagens in tobacco cause the sort of DNA damage that results in C to A mutations. In other words, the mutations that were being found in the cancers bore the imprint—what we today call the signature—of the exposures that caused them.

Today, we have the genome sequences of somewhere between 25 and 30 thousand cancers. With these, we can now reexamine this question to see if we can get a comprehensive view of the different patterns, the different signatures of mutations, that are present across the face of human cancer, and let that give us a sense of what the different processes are that are causing the mutations that cause cancer.

Elizabeth McKenna: What are some of the other signatures you’ve identified and what are they associated with?

Sir Michael: Thus far, about 40 different signatures have been found. That’s probably more than we were collectively expecting; 40 signatures basically means 40 different mutational processes. As we analyze more cancer cases, the number is increasing. It’ll probably hit around 50 or perhaps a bit more. About half of those, we actually have no idea what causes them at all. Basically, half the causes of cancer are, at the moment, opaque to us.

We do have some idea about the other half, or at least some suspicions and in some case, more than suspicions. One of the major findings in the last few years has been the discovery of signatures that are attributable to the activities of the so-called APOBEC enzymes. The normal, day-to-day job of this family of enzymes is to edit DNA. APOBEC enzymes are cytidine deaminases. They remove amine groups from cytosine and convert cytosine to uracil. If DNA replication takes place when there’s uracil in the DNA, it gets converted to a thymine, so you end up with a C to T mutation.

These DNA editing enzymes have a normal day job. They do good things for us. They are believed to be involved in innate immunity and protecting us against invaders. For example, when viruses invade cells, the cell senses it. One of the ways that it can protect itself against viruses or other forms of naked DNA is to mutate them, to essentially “machine gun” them with mutations, such that the virus can no longer function. That is one of the functions of this group of APOBEC enzymes.

In many cancer types, we find cases with very large numbers of mutations that we believe are caused by these APOBEC enzymes. There are two signatures that we believe are the result of the activity of APOBEC enzymes.
About a third of breast cancers have got a lot of APOBEC mutations, as do almost all bladder cancers, whereas we rarely find them in colorectal cancers. You’ve got this enormous diversity of exposure to this mutational process that is absolutely overwhelming in the number of mutations that it generates when it happens.

The other mystery is, what is switching on the APOBEC enzymes to create this tidal wave of mutagenesis in the cancers in which it attacks? Here you end up essentially “hand waving,” because we don’t know the answer. We believe it has to do with the activity of the APOBECs. Perhaps some time in the past, decades ago, a virus—any old virus, really—entered the cell and the cell decided to protect itself against that virus. It decided to mutate the virus, but at the same time, it couldn’t control its activities and when it deployed the APOBEC against the virus, there was collateral damage on the nuclear DNA as well. That cell would be sitting there for decades, but now with many more mutations than the cells around it. Those mutations would mean that cell was primed to become a cancer cell compared to those around it.

If that idea is true, it gives us a cause of cancer we hadn’t anticipated previously: That a variety of different infections that otherwise we had not linked to cancer could be doing this. As a hypothesis, it feels plausible, and it certainly would be interesting. The trouble is, we don’t have much evidence in favor of this. For example, we don’t find great numbers of APOBEC mutations in liver cancers where hepatitis B and C viruses have been operating and you might have expected them. You do, however, find a lot of APOBEC mutations in almost all cancers of the cervix, and in a subset of cancers of the head and neck. In both those cases, human papilloma virus is known to play a role. You can’t dismiss the hypothesis, but at the moment, we also can’t prove it.

Elizabeth McKenna: What about other commonly mutated genes, like driver genes? Are they associated with specific signatures?

Sir Michael: Some are. The best examples are the breast cancer susceptibility genes BRCA1 and BRCA2. It’s been known for a while that when BRCA1 and BRCA2 are inactivated by their mutations, the cell is prone to suffering rearrangements of DNA that are aberrantly repaired, such that one bit of DNA is joined to another bit of DNA when it shouldn’t be, and that can inactivate or activate a cancer gene. The mutational signature analyses showed that it’s not just these gross rearrangements. There is a substitution mutational signature that is absolutely clearly associated with defective BRCA1 and BRCA2 function. That gives us new insight into the biology of BRCA1 and BRCA2 deficiency and how the susceptibility is orchestrated. In the future, that signature could also be a good predictor of patients who will be treatable with particular drugs. It’s already known that BRCA1 and BRCA2 mutant cancers have a deficiency in homologous recombination-based repair. Those cancers are vulnerable to DNA damaging agents like cisplatin and PARP inhibitors. Currently, those patients are identified on the basis of having either a BRCA1 or a BRCA2 mutation. However, what is clear from the signature analysis is that although both those groups do have this signature (which we call Signature 3), there is another subset of patients of more-or-less equal numbers that clearly have the signature and have the abnormalities of genome rearrangement too, but you can’t find the BRCA1 or BRCA2 mutations. If this is correct, then the signature is identifying much more widely the cancers that have a deficiency in homologous recombination repair and therefore could identify with greater sensitivity those patients that are going to be responsive to PARP inhibitors and cisplatin and other drugs. That’s to be seen, but I would be optimistic that the signatures will be used as predictors for therapeutic deployment.

Elizabeth McKenna: You envision the signatures being used for patient stratification and things like that?

Sir Michael: Absolutely. It’s not so straightforward to get the signatures. Small segments of genome being sequenced don’t give you quite as much information as you might need. We might need at least exomes, but possibly whole-genome sequences to be absolutely sure of the presence of a signature. I’m optimistic that, over a period of 5 to 10 years, whole-genome sequencing of cancers to find these signatures associated with DNA repair deficiency will be used to find the patients who are going to be responsive to those drugs.

Elizabeth McKenna: Are there signatures associated with aging, or is it mainly just mutations and things you’re exposed to?

Sir Michael: That’s an interesting question. There are two signatures that we call 1 and 5. For both, the number of mutations that you find in a cell correlates with the age of the person. These signatures reflect underlying mutational processes which are essentially operating in all of our cells throughout the course of our lives. If one is a 20-yr-old, one has had 20 years for one’s cells to acquire mutations of these two signatures, so one will have a certain number. If one is a 40-yr-old, then one will have double the number of mutations in cells of one’s body. Both of these mutational signatures are mutational clocks. They are slightly different from each other, however.

Mutational Signature 1 is a very consistent clock; it can’t be deflected. We believe that Signature 1 reflects the number of mitoses a cell has been through over a lifetime. We have cells in the tissues of our body that are turning over, day after day, year after year, decade after decade. Each time a cell divides, it goes through DNA replication and a small number of mutations are acquired. Those seem to be Signature 1 mutations. In the future, actually finding the number of Signature 1 mutations will give us great insight into how many divisions a cell has been through. These mutations add up throughout life. If they fall in the wrong genes, they can convert a normal cell into a cancer cell.

Signature 5 is more mysterious. It is also clock-like, and the rates at which it generates mutations are different
in different cell types but it differs from the rates at which Signature 1 produces mutations and is not correlated with the number of mitoses a cell has been through. Frankly, we have very little idea what is causing it. We think it can be deflected from keeping time, and there are various things that can cause a cell to acquire more Signature 5 mutations. We believe smoking is one of those things, and there probably are others. Signature 5 is clock-like, but it’s a floppy clock: Other factors can push up the rate. I don’t know if you can push the rate down.

So, yes, we do have the seeds of our destruction within us. We are acquiring mutations all the time. These are the two major processes that we think are doing so. Indeed, we believe that these two processes, which are operating in all the somatic cells in our bodies, also operate in the germline to give all the mutations that create the differences between us: the phenotypic differences in the way we look and behave and our predispositions to disease. Remarkably, although we do have an idea what is causing it for Signature 1, for Signature 5 (which is about two-thirds of the mutations that occur in a germline), we really have no idea. There’s all this interindividual variation, and at the root of it is this mysterious process, and we need to find out what is doing it.