

This Week in Microbiology

With Vincent Racaniello and Sam Sternberg

Episode 184: CRISPR-Cas immune systems

Aired September 7, 2018

<http://www.microbe.tv/twim/twim-184/>

Vincent: This Week in Microbiology is brought to you by the American Society for Microbiology at asm.org/twim.

(music)

Vincent: This is TWIM, This Week in Microbiology, episode 184, recorded on August 9th, 2018. I'm Vincent Racaniello, and you are listening to the podcast that explores unseen life on Earth. Today my guest is a pretty recently minted professor here at Columbia University, Irving Medical Center, Sam Sternberg. Welcome to TWIM.

Sam: Thanks, Vincent.

Vincent: Newly minted, is that a good descriptor?

Sam: I just had my 6 month lab anniversary, so it's been just over 6 months, yeah.

Vincent: That's pretty good. I've been here, let's see, August and September I will have been here 36 years.

Sam: Alright, well, hopefully I'll be sitting on the opposite end of the table at some point and I can be having this conversation with the next generation.

Vincent: (laughs) You may not be here for 36 years, but you'll be somewhere. I just never left. But some people move all the time, you know. Just how it is.

Sam: I did my undergrad here, so I'm coming back home and I wouldn't mind staying here for some time. We'll see.

Vincent: So you were an undergrad at Columbia and that means down at the 116 Street campus, where are you from originally?

Sam: Pennsylvania, Lancaster, Pennsylvania.

Vincent: Lancaster, now what is there, is there a train there or something like that?

Sam: Amtrak runs through there but it is best known for Amish country.

Vincent: Amish country, right.

Sam: So people think oh, you must be a farmer, you must be Amish. I grew up in Lancaster City, it's not a big city but you know, you see Amish at the farmer's market downtown, which is fantastic, but the day to day life there, they are out in the countryside. But beautiful farm country, good city, good place to grow up.

Vincent: Did you get interested in science from a young age?

Sam: I did science fair projects, my dad was a college professor in geology and geophysics, but my science fair projects were mostly in the Earth sciences, which I think as a high school student, middle school student, didn't really inspire me. So I really have to credit Columbia and a couple of classes in my first two years that really got me on the first organic chemistry pursuit and from that I moved more to biochemistry. I remember taking biochem with Brent Stockwell, and Li-Yang Tan, and loved drawing out mechanisms, thinking about molecular details of how chemistry happens in the cell, so I think my real passion for research and for science really started in college and then especially working with my undergrad adviser, Ruben Gonzales, that was what kind of told me this is what I want to do in life.

Vincent: What years were you at Columbia?

Sam: 03-07.

Vincent: So I hadn't started teaching my virology course, I think I started the next year. You probably would have taken it.

Sam: I would have loved to, I have to confess I don't know much about eukaryotic virology. I think phages are pretty cool, we're gonna talk about that I'm sure. Since grad school cloistered in the worlds of bacterial and bacterial viruses more than eukaryotic viruses.

Vincent: We talk a little bit about phages in my course but since I spent my life working on eukaryotic viruses, that's what dominates the course. So you didn't come to Columbia pre med, then?

Sam: Nope, I was, I didn't really know what I wanted to do but I knew probably math or science, something in that area. Flirted a little bit with pre med at some point but for me, I love being in the lab, I love thinking about basic mechanisms, I'm happy to be at the medical center now, because I think being pushed a little more in the direction of thinking about the translational science is always a good thing. At the end of the day I like thinking about how things work in the cell and so for me, pre med was never gonna satisfy that curiosity.

Vincent: Where'd you go after here?

Sam: So I went to UC Berkeley for grad school, I actually tech'ed for a year and a half at Columbia in the same undergrad lab, Ruben Gonzales, then I did my PhD at UC Berkeley and that is where I got into CRISPR before CRISPR was a huge fad, and then I actually worked at a biotech company for a year after writing a book with Jennifer on the discovery of CRISPR immune systems and the development of gene editing technology using CRISPR, so I had a brief stint as a book author then as a scientist in industry and then I came back to academia and I'm exactly where I want to be.

Vincent: Did you go to Berkeley to work with Jennifer or?

Sam: She was probably top 2. The drama of my grad school beginnings was that I deferred last minute for a year, and then in that year of deferral, I both found out that she had moved to Genentech to take on a vice president position and also that was around the time of the great recession and Berkeley happened to be the only state funded university that I had applied to. All the other grad schools I considered and said no to were private, and California was hit amongst the hardest by that recession, so I got an email from the university having deferred already that they were doing campus wide 20% budget cuts, cutting back of staff, custodial staff, so those were kind of a bit of turmoil before I actually started. But then Jennifer ended up resigning from her position before the lab even relocated, came back to Berkeley. I started, did a rotation at her lab, was convinced that she was happy to stay at Berkeley long term, and then yeah. I was very happy spending my PhD in the lab there.

Vincent: I didn't even know she had gone to Genentech for a while.

Sam: Yeah, I think it surprised a lot of people. The lab, you know, there were some casualties of that brief move, some people left the lab because they didn't want to finish their PhD or do a post doc at company, but what she told me is she realized unfortunately after having taken the job and she herself was already spending 4 days a week there, but she just realized very quickly it wasn't actually the place she wanted to be. And I think certainly it benefited the lab to stay in Berkeley because that I think is just a great environment. The building we were in has a lot of structural folks, a lot of people doing biochemistry and molecular biology, so for me as a student that was a fantastic place to study.

Vincent: What building were you in?

Sam: Gosh, what was it called?

Vincent: It wasn't Lee Kashing was it?

Sam: No that wasn't there yet. Wow. Stanley Hall, which used to exist. Before the new building it was knocked down before I started, and then the new Stanley Hall was built but yeah, Li Ka Shing was going up as I was doing my studies and a couple other buildings down in that part of campus.

Vincent: So I went out there a couple of years ago to do a podcast. They have a student run symposium every year, I think it is the micro department but I'm not sure. I came and did a podcast. It was in the Li Ka Shing building and we had a dinner, so the labs have terraces. And they overlook the bay. We had dinner one night out there, it was just great. Someone took me to the museum to show me the T-Rex skeleton there, which was just awesome. Great. You must have gone to see that, right?

Sam: Of course. The Bay area view will spoil you forever. I remember coming back to York and seeing an apartment with a Hudson River view and it was so unspectacular because compared to the Bay area view it is just a brown looking river. I enjoy looking at that every time I see it but the Bay area is something else.

Vincent: It's very nice, yeah, it's very nice. So when you joined the lab, had CRISPR already been taken into the lab at that time?

Sam: So when I joined, my rotation project was actually studying micro RNA biogenesis by human DICER, and by that time about half the lab was studying RNAi and CRISPR was being pursued by one post doc and one grad student who had joined a year and a half before I did my rotation, but it was really like an orphan project that the lab didn't really care about and most people didn't understand, so it was interesting and it appealed to me because compared to RNAi you could spend one week and read the entire literature on CRISPR. Literally every paper on CRISPR both after it was coined and even all the papers before the term itself was coined. You could read it all in one week.

For me that was actually very appealing as a research topic to be in a field where there wasn't a lot of competition. There were a lot of open questions. No one really knew how anything worked beyond the broad strokes. At that point it was known that these were immune systems widespread in bacteria and archaea, but in terms of what the molecular components were, how the RNA was being used by different proteins, that was a big black box.

Vincent: Of course, nowadays you can't read the literature in a week.

Sam: No. Nowadays you can spend a week trying to catch up on just the papers that came out in that week.

Vincent: That's amazing. I did a search just for reviews and they are so specialized. Some of them are talking about using CRISPR in specific organs. Amazing. Just amazing.

Sam: Or animals or plants.

Vincent: Yeah, pretty cool.

Sam: I struggle nowadays with just staying on top of the literature in deciding which papers should I be reading because yeah, it's just the volume of work coming out was pretty staggering.

Vincent: It's kind of interesting to be at the genesis of a new field. Have you ever thought about that? You were at the beginning, really, and now you are seeing it exploding. Doesn't happen very often, right?

Sam: Yeah. You hear older professors talk about how something began. It's fun that I will now be able to talk about CRISPR in that way, not that I played any significant role in that. But you know, I was at the second annual CRISPR meeting ever and this is the meeting that is still going on every year, but now it's one of a few dozen meetings every year, but this is the meeting that began in 2008 focused purely on bacterial and archaeal adaptive immune systems, and so, those were the meetings where all the people in the world studying CRISPR were coming to Berkeley and it was an audience of maybe 30-40 people. That was the world's CRISPR community back in those years.

Vincent: So when I started as a PhD student, recombinant DNA had been developed, which was also an enabling technology, right? It was amazing. But we didn't quite realize it at the time, I think. It takes

time for things to accelerate. Over the years there have been other cool things. I think PCR was a big one. It has permeated everything. The cool thing about science is it's always something new, right?

Sam: Absolutely.

Vincent: People just are curious and they'll find new things. I think that's the cool thing about it. I wonder if you could give us a brief history of CRISPR. How did it start and bring it to the present, doesn't have to take a long time but I'm sure you could give us an overview.

Sam: So I often, when I give talks for non specialist audiences, I always show a screenshot of the first paper describing what we now call CRISPR. That was in 1987. This was back when sequencing genes was an entire publication, so they were sequencing a gene in *E. coli* and in the 3' UTR of that gene they found a series of direct repeats that were spaced in this very tandem array where the same 32 nucleotides of space of DNA were separated each repeat. So that was the first time this feature was described.

Vincent: They had no idea what it was, though, of course.

Sam: So they have this classic sentence in the very end of the paper that the biological significance of these sequences is not known. And as I comment on in my talks, that was going to be the case for another 20 years. Of course, what was different 20 years later was that these had now been found, not the same sequences of repeats, but the same repeat properties, namely direct repeats, interspaced with other sequences. Those had been found in about 40% of all bacterial genomes that have been sequenced, something like 90 or 95% of all archaeal genomes.

So by the early 2000s it was appreciated that whatever these were doing, they were extremely widespread and that is of course a great indicator that they must have some important function, because they wouldn't be retained over evolutionary time if they are not doing something critical to the species. Then in 2005 were the first clues about what their actual function might be. That was a result of 3 independent papers that found if you ignored the direct repeats and you looked at the sequence of DNA spliced in between the repeats, those were often perfect matches to other bacterial virus DNA or plasmid DNA.

So foreign genetic elements that are known to parasitize various bacterial or archaeal hosts. So that led to the very tantalizing speculation that maybe these sequences were being stored in the host genome as a way to serve as some immune system, some sequence feature that would help identify a foreign genetic element as being foreign. So that was still speculation at that point, and then I think the real breakthrough paper that really put CRISPR on the map was a 2007 study from a yogurt company at the time called Danisco, they were studying the bacteria *Streptococcus thermophilus*, which is the main workhorse bacterium to ferment milk into yogurt, various cheeses. Obviously they have a major interest in developing more virus resistant strains because those are going to save them a lot of money in the long term.

They were the first to actually take viral, what are called bacteriophage insensitive mutants, and we should probably mention that bacteriophages is just a term for bacterial viruses, phage meaning to eat. So they were isolating colonies that were resistant to viruses during experiments where they were

infecting them with different viruses, and then they actually did DNA sequencing of the CRISPR part of the genome, and they found that these strains that were acquiring resistance were actually expanding their CRISPR DNA. They were splicing new sequences from those viruses that had been infected with into the CRISPR array and those strains that had new sequences in the CRISPR array were now immune to any phage with a matching sequence. So that was really the first study that proved that these sequences provide adaptive immunity.

Vincent: They didn't remove them though and show that they reacquired susceptibility, right?

Sam: They didn't go that direction, but they put them in and then infected them. They had one less experiment where they had two different phages with partial homology so they could show that one new spacer that was acquired from one phage provided resistance to a different phage that had the same target sequence matching that new spacer.

Vincent: This brings up a question that I always had. So you have, these are lytic phages, right?

Sam: That they were using, yeah.

Vincent: So if infection kills all the bacteria, how do you get some that have this phage DNA incorporated into them? Are there just a few survivors that then go on and be resistant?

Sam: So in the context of acquiring immunity, they are just putting a little snippet of the phage DNA into the CRISPR array. But it is an interesting question because actually one of the discoveries that really shocked the field in 2012, 2011, I think it was 2012, was published by a friend of mine at UCSF, Joe Bondy-Denomy, and he asked a question, why do some phages escape the CRISPR system? Why can some phages lysogenize the host and go undestroyed? And it turns out that those phages have phage encoded inhibitors of CRISPR systems, or anti-CRISPRs as they are called. So it's been really fun being in the field for so long, you've kind of seen those first discoveries, just tell us on a broad level what is going on, but as the years have advanced we've learned more and more about all of these interesting features of this arms race between bacteria and virus.

Vincent: So yogurt was the first demonstration that these are probably involved in immunity to infections, phage infections.

Sam: That's right. I had the fortune of meeting the scientist that coined the term CRISPR. So that came out in 2002. But I like to mention him, this is Ruud Jansen who is in the Netherlands, because not only did he coin CRISPR in his paper in 2002, but he also made a really important discovery, which is that these CRISPR arrays are consistently flanked by a core set of conserved genes which we now call CAS or CRISPR associated genes.

So by 2007 it was known not just that these CRISPR arrays have sequences matching virus or matching plasmids, but also that they are probably working together with proteins encoded by genes that always flank the CRISPR array. So in 2007, that study also showed genetically that some of those genes were required for this resistance phenotype, but it would take another number of years to begin piecing apart what each of those protein products are actually doing.

So now we know some of them help process RNA transcripts encoded by the CRISPR array into these guide RNAs, some of them form targeting complexes, CAS9 of course is the very famous example, but there are actually many other different kinds of targeting complexes using different protein components. Then there is the nearly universally conserved gene called CAS1 which produces an integrase protein that is required for splicing new sequences into the CRISPR array to provide this adaptive immunity.

So the immune system is often broken down into three stages. One is integrating new sequences into the CRISPR array, so splicing in that new sequence during the vaccination stage of the immune system, if you will. Then the second stage is producing the guide RNAs and all the protein products and assembling the targeting complexes. And then the interferon stage is when all these components come together and destroy complementary nucleic acids that would be encountered during a reinfection by that virus.

Vincent: So the realization was that these are both phage or plasmid sequences that get incorporated. As you know, horizontal gene transfer involves plasmid DNA going from bacterium to bacterium and that can be good, right? So how does the system identify good versus bad plasmid DNA, or maybe it doesn't, I don't know.

Sam: It doesn't, really, but it's not 100% effective, so I think, that's the interesting thing from the perspective of the host. You want to avoid allowing phages to destroy a population, but exactly, some horizontal gene transfer also phages can be very beneficial to the host. I think it is really a balance of those two pros and cons.

Vincent: So why is the, you mentioned CAS9 is the most famous, why is that?

Sam: Well, gene editing.

Vincent: Because it's been adapted for gene editing.

Sam: By most famous I just mean today in biotechnology. You often hear CRISPR used synonymously with CAS9. In fact, there are some sticklers that would say using the term CRISPR for gene editing is actually a misnomer because there is no CRISPR array in the tool. You are using one protein and a synthetic version of a guide RNA. That's not a CRISPR either. But I like to point out often that CAS9 and the type of CRISPR system it comes from, that's actually one of the less common CRISPR systems that exist in nature.

It so happens that of course it is extremely effective as a tool because it is a single protein. You can make these single guide RNAs, you can transfect CAS9 in the guide RNA very easily into many different cell types, but the biodiversity of CRISPR-CAS systems in bacteria and archaea, CAS9 is a minority player and one of the things that my lab is thinking about is all these other flavors of CRISPR-CAS systems that could be potential gold mines for tool development, but I think also coming back to the basic research questions, how do bacteria use totally different protein and RNA architectures to accomplish the same thing, which is recognizing foreign DNA in a very specific fashion and neutralizing it.

Vincent: So let's cover first how this works. Let's say we have a bacterium with a CRISPR array and it is infected by a phage whose DNA is encoded in that array. What happens? Are these CRISPR RNAs always made or are they induced?

Sam: That's a good question. It depends on the system. Ten years ago, or how many years ago, eight years ago or so, *E. coli* was one of the model systems being used to study CRISPR. But in *E. coli*, lab strains of *E. coli*, the CAS genes are transcriptionally repressed by a global transcriptional repressor, and *E. coli* has no functioning CRISPR system even though the genes are there and the CRISPR array are there. If you move those on to plasmids and over express them artificially, you can now develop resistance against viruses or plasmids.

But for whatever reason, those *E. coli* strains don't function in their wild type state for immune defense. It's a really neat paper from Konstantin Severinov who is at Rutgers and also has a lab in Russia. They actually sequenced *E. coli* from a woolly mammoth carcass which had been preserved for 40,000 years, and showed that the spacers within the CRISPR array in those *E. coli* strains were virtually unchanged compared to present day.

So that gives you a sense that for whatever reason that system has no longer functioned as an adaptive immune system in *E. coli*. But it works if you artificially over express it. So there are other systems like *Streptococcus thermophilus* where I'm sure there's been RNA seq done. And there I believe these are being constitutively produced, you're always making guide RNAs, you're always making CAS9 so that you have this supply of surveillance complexes that are always patrolling the cell looking for foreign DNA.

Vincent: So it must be a good fraction of the total transcriptome, right?

Sam: The RNA, absolutely, and in fact that was one of the ways that this other type of RNA was discovered, which ended up being a critical enabling discovery for the use of CAS9 for gene editing. Some of your listeners will have heard of tracer RNA, so in some CRISPR systems there is both the guide RNA encoded by the CRISPR array, that's known as CRISPR RNA, but there is a trans activating CRISPR RNA or tracer RNA that is expressed outside of the CRISPR array itself, and that was actually detected because if you do RNAseq on the bug where it came from, *Streptococcus pyogenes*, the tracer RNA is one of the most abundant small RNAs in that bacterium.

Vincent: So is the entire array, I guess these arrays can be pretty big, right? Is the whole thing transcribed as one transcript or multiple transcripts?

Sam: All transcribed as one transcript. You generally have lower abundance of spacers within CRISPR RNAs at the far end of the transcript, because you can imagine getting RNA polymerase drop off as you extend down the array. And that could be beneficial because the neat thing about the integration of new sequences is that happens directionally. So there is always the newest spacers at one end of the array that are transcribed first. And you can imagine for a bug, you are storing memories from past infections in your CRISPR array. You probably don't care that much about the spacers you got a million generations ago, because the likelihood of encountering viruses that you encountered a million generations ago is probably quite low. You care about having those spacers from the most recent

infections at the part of the array that is going to be transcribed first and have the highest abundance once those RNAs get processed.

Vincent: Do we understand how those newest infections are integrated at that one spot versus the other end, right?

Sam: So that's something that is pretty well understood now. That was the last piece of the puzzle to get filled in in terms of the basic mechanisms of each of those three stages. A friend of mine in Jennifer's lab who is now also at UCSF is a post doc, he did some beautiful structural work, looking at structures of the integration complex, showing how sequences are recognized within the repeat but then also host factor that is involved in bending the DNA in a very characteristic way just upstream of that first repeat. That helps localize this complex to that first repeat and also recognizes the sequences in front of it which are called the leader sequence. So it is pretty well understood now how you get very targeted integration only at one position within the array.

Vincent: So how long is the transcript? I guess it depends on experience.

Sam: It can be really long and it is very fun gazing at CRISPRs. I'll go back to this Dutch scientist who coined CRISPR. So he actually showed me his notebooks from the late 90s back before you were doing all your sequence analysis on a computer. So he had these printouts, a fat binder, he had printouts of a genomic sequence with all the repeats highlighted in yellow. It's really fun to look at these because of their geometric nature. You have this beautiful kind of periodicity that really jumps out at you, even without highlighting the repeats. If you kind of blur your eyes a little bit, you can just see patterns where the same letters are cascading across the page in the same pattern.

So he, and some of these arrays are hundreds of spacers long, some are quite short. Those shorter ones are not necessarily any less effective. There is thought to be some turnover of old spacers through recombination. You could imagine because these are direct repeats, they are gonna be pretty good at recombining. The idea is that probably there is some equilibrium of acquiring new spacers over evolutionary time, losing old spacers to recombination. But the idea is you've got transcription across an entire array, and then there are dedicated ribonucleases that will recognize sequences within the repeat portion, and introduce a cut in the RNA so that from this long single transcript you get a library of short, mature CRISPR RNAs.

Vincent: Some of those will recognize incoming phage, and what is the sequence of events, what happens next? Does that CRISPR RNA hybridize, does it bind something else, what happens?

Sam: So part of my PhD research was understanding this process of target search. How do these proteins sift through the vast expanse of not only potential incoming phage DNA but they are not gonna be able to discriminate a priori against phage DNA or genomic DNA, so they've gotta look through megabases of DNA, looking for that one target. So we did a bunch of work trying to understand how they can pare down the search complexity of the genome and the sequence motif called PAM, or protospacer adjacent motif, it is a short 2 to 5 or 6 letter code, depending on which CAS9 or which targeting complex you are studying. But these are used as a hot spot so that you only invest energy looking for complementarity in the DNA. If there is this short motif flanking the DNA sequence.

So basically these complexes pop around using just 3D diffusion. Every time they hit a PAM, they start unwinding the DNA a little bit looking for a potential match and at a cognate sequence, the DNA can be fully unwound. And then there's some proofreading checkpoints that avoid premature cleavage of the target DNA unless it is really the right match. In the case of CAS9, it makes a double strand break by cleaving both strands using two different nuclease domains.

Again, I think it is fun to point out that is one way that CRISPR systems target DNA. There are other enzymes that cleave DNA in a completely different way, instead of making one double strand break. They actually unwind and processively cleave at many positions using a protein that is a helicase and a nuclease all in one, and then there are whole types of CRISPR systems that don't target DNA at all, they target RNA. So there is actually, again this vast diversity of different ways that different systems have evolved to use the same guide RNA hallmark feature, but use it to recognize foreign genetic elements in totally different ways.

Vincent: And there are probably more that we haven't found, that's why you want to keep looking.

Sam: Absolutely, absolutely.

Vincent: So the CRISPR RNA is searching. Does that happen together with CAS9, already attached via the tracer RNA, so tracer RNA, how are those RNAs brought to CAS9?

Sam: That's not super well understood. Presumably CAS9 just bumps into CRISPR RNA, tracer RNA, those are already hybridized. So those have a large region of complementarity, so the tracer RNA and the CRISPR RNA can hybridize through the stem that forms between them. That's actually critical for the processing of those long transcripts. So in the type II system that uses CAS9, there is a different kind of processing mechanism that relies on tracer RNA and CRISPR RNA hybridizing. So those hybridize. That leads to cleavage by a host factor called RNase 3, and then you have the right substrate for CAS9 to bind to and it has sequence specificity for that particular type of hybrid RNA. So one of the other things is because you have the CAS genes like CAS9 directly flanking the CRISPR array, you could imagine that you are getting localized transcription of the CRISPR array, transcription and translation of the proteins, so these are probably all spatially confined and so you have a much higher probability of these things interacting with each other.

Vincent: So in an uninfected cell, these are being made unless you are E. coli, in which case, right. But so why don't they cleave genomic DNA, because they have the CRISPR arrays in there?

Sam: So that's another part where the Pam guided target search becomes key, because the CRISPR array has the spacers which, as you already pointed out, are a perfect match to the guide RNA because that is where they are encoded. But the sequence flanking the spacer in the CRISPR array does not have a PAM. And so that is an immediate way that CAS9 will always exclude the CRISPR array itself as a potential target, because the absence of a PAM means it is not even gonna ever look for a potential match there.

So the PAM is really the first requirement that even allows it to begin melting open the DNA. So we think that it probably evolved as a requirement to avoid self targeting, which any immune system

needs to be able to do, avoid self and discriminate between self and non self. And then it also serves as a way to reduce search complexity, because instead of having to pry apart the DNA double helix every time you bind, you just invest that energy when you are at a PAM, which represents a potential target site.

Vincent: The CAS9 is not active unless these RNAs are bound to it, right? Do we understand how the binding triggers activation? I guess it is DNase activation.

Sam: Did you read my papers? I feel like you're asking me questions that are--

Vincent: I did, I looked over all of them. I didn't read every one but I looked at them.

Sam: I'm now realizing that these questions are a little bit targeted to some of the work I've done.

Vincent: But for our listeners it is a good way for them to focus, because you've touched on a lot of things that impact the entire field, so. Anyway, it's your interview. So.

Sam: I actually love talking about the CRISPR things that, I'm trying to expand beyond, my PhD was also focused on very nitty gritty mechanistic questions. I think what I liked about these projects is understanding target search is interesting as a basic research question, but that actually becomes critical when you think about how do you use this tool and avoid off target effects, and that is another reason why on the most recent question we were very interested in understanding why does CAS9 not cut DNA sequences that it has a high degree of binding affinity for? Because we know from a variety of experiments, including work that was done by a new professor starting this fall, who was one of the first to do CHIP-SEQ on mammalian cells expressing CAS9 and guide RNA, and he found that actually CAS9 will pull down a lot of mismatched DNA sequences that have very limited homology to the guide RNA.

So that really begs the question, why is CAS9, and if you do a Venn diagram comparing sites that it binds to and sites that it introduces double strand breaks in, it is much much more promiscuous in binding than it is in cutting. That really begs the question, how does this protein have greater DNA cleavage specificity than the binding specificity, which is actually quite low? And there were some very interesting clues from the first crystal structures that came out.

So being in Jennifer's lab, I didn't actually do any structural biology myself, I did collaboration, but, you know. I spent a lot of time thinking about how structures could tell us about mechanism, and in fact, if you looked at those first structures, the nuclease domains of CAS9 which actually cut the DNA were bizarrely mispositioned. They were very far away from the actual cut site on the DNA, and there wasn't a clear explanation of why they were out of place. My hypothesis was maybe the positioning of the cutting domains is actually being controlled to improve the specificity beyond binding. You might initially bind DNA in some inactive state, and only with sufficient complementarity would these domains undergo some conformational change that would then trigger cleavage.

So that is exactly what it turns out to happen, and we did some fluorescence based experiments to show that there is this conformational control, and in fact if you now look at a number of the papers that have been published on higher fidelity CAS9 variants, again around the technology space where

people care about off target editing and have been using a variety of methods to look at off target effects and then use directed evolution or rational engineering to make higher fidelity variants. We published a paper last year where we used this knowledge about conformational control to build a higher fidelity version, and there were two other studies that didn't have access to that mechanism yet because we hadn't published our paper, but if you now look at where their mutations are found, it works together beautifully with our conformational control study, because those are directly in those positions that would affect this conformational equilibrium that we think CAS9 has evolved to control.

Vincent: So what is the extent of base pairing between the RNA and the DNA, how many bases is that?

Sam: 20 base pairs in total, and we actually focused on a small number of model mismatch substrates to show that if you have 17/20 that are correct, that is kind of the right, when you cross over between undergoing this conformational change and cleaving or not cleaving, and that actually matches well with other off target studies. But this has obviously been done in a much more high throughput scale with some of the studies looking at editing efficiency.

Vincent: So three bases mismatch is enough for it to hit somewhere else. Say you're doing gene editing, it could hit somewhere else in the genome, which would be bad. Right? So you'd like to get around that obviously.

Sam: Yeah. And I would say for gene editing, you would like to be able to discriminate one base per mismatch, too. If you're talking about using this therapeutically, you know, you really want it to be as close to perfect as possible. Some of these higher fidelity variants have improved specificity, where in many cases it can discriminate just a single base pair mismatch, which is really important if you are gonna move this into the clinic.

Vincent: So it will, you can't get 100% fidelity, it will only recognize 100% of those 20 base pairs, is that possible?

Sam: Depending on the target sequence, there are some targets where people have looked and found zero off target effect cleavages elsewhere in the genome. Some of the algorithms now for choosing guides, if you can choose a guide that has very few similar sequences elsewhere in the genome, then you may not even have any off target potentials that are just different by one base pair. It's a combination of just bioinformatic predictions of how many sequences are similar and then doing the experiments to look in an unbiased fashion if you get any edits at off target effects. There are a number of studies now on how you can do this unbiased analysis.

Vincent: Let's talk a little bit about the finding. You mentioned that diffusion governs the ability of the CAS RNA complex to find its target. So what is the time scale that we are talking about? Especially a eukaryotic genome, but even an archaeal or bacterial genome is pretty big. Are we talking about milliseconds to scan the entire genome or longer?

Sam: More like hours.

Vincent: Hours! (laughs)

Sam: Sorry, hours for a single CAS9.

Vincent: There's probably more than one, right.

Sam: For it to be effective against a phage, you need to do it probably in about 15-20 minutes. The reason hours was in my head is because there have now been some studies looking at search kinetics using fluorescent labeled versions of CAS9 in E. coli. People have also looked at target search in mammalian cells, and the results of the study on E. coli found that for a single CAS9, it takes somewhere on the scale of I think a couple hours, which isn't obviously a relevant number because E. coli is dividing many times in those couple hours. But obviously the time to find a target is gonna depend not just on search time per molecule but how many molecules are undergoing the search, and that also comes back to how big is the CRISPR array and how many different guide RNAs are soaking up CAS9s, because now you gotta look at what population of CAS9s have the guide RNA that match the incoming phage.

Vincent: What is actually happening? The RNA protein complex, is it actually scanning the entire DNA base by base?

Sam: We think it is just diffusing three dimensionally. So I came to Columbia during my PhD to work with a professor named Eric Green to use single molecule fluorescence microscopy looking for scanning. We thought maybe these use one dimensional diffusion that they might kind of bump into the DNA and then slide one dimensionally along the DNA double helix. That might be a way that they accelerate search time, and we didn't actually find any evidence for those 1D diffusion dynamics. So we think it is mostly popping on and off DNA through random collisions and you just need enough CAS9 molecules and enough time to happen to find the right spot.

Vincent: That's amazing.

Sam: That works because you think yeah, if you're searching through a megabase per genome or now talking about a human cell, that has a gigabase per genome, for this to actually work is kind of amazing.

Vincent: I was gonna get to that, let's do that a little bit after we talk about gene editing. Because we have chromatin there and everything, right.

Sam: That's another, yeah.

Vincent: So, alright. It's bumping randomly, there are lots of CAS9 RNA complexes, eventually it sees some complementarity. The PAM is there and that triggers conformational change, double stranded DNA break in the case of CAS9, and that wrecks the virus, basically. Is that a fair summary?

Sam: Yeah. You know, one thing we found many years ago is that unsuspectingly, CAS9 stays bound to the cleaved DNA. It, in a test tube experiment, for quite a long time, and this has now been shown in a bunch of other labs and in fact a paper just came out recently showing that in gene editing experiments, you can accelerate the editing efficiency if you have your target in an actively transcribed region because rather than CAS9 staying bound, if RNA polymerase, POLII, knocks it off the DNA, that

can actually affect repair in a way that you can imagine now you've exposed the double strand break to host factors and the quicker CAS9 gets ripped off, the faster that you might have repair factors coming to the DNA.

So I've always actually wondered, what's the relevance of CAS9 staying bound to cleave DNA in the context of a phage infection? It's not clear to me that it would be beneficial to cut the DNA but then hold on to the DNA. I always thought maybe if you don't protect that double strand break, maybe the phages can recombine if you have a multiplicity of infection above one and you just repair the double strand break with a neighboring phage genome, whereas if you hold on to it until a DNA polymerase comes along and then you get some messed up replication fork because you've got a double strand break, maybe that is actually better than just cutting once, letting go, and giving the cell the chance to homologously recombine two different phage genomes.

Vincent: It makes sense to me, because yeah, you're right, recombination could fix it for sure.

Sam: So we think it cuts and it may stay bound and at some point, that phage genome can't be properly replicated anymore and that is sufficient to ablate the infection.

Vincent: So let's talk now about how you modify this for gene editing. That was, was that done while you were there in the laboratory?

Sam: So actually I didn't start working on CAS9 until after Martin Jinek, who is a post doc at the time, started this collaboration with Emmanuel Charpentier, she was the one whose lab had discovered tracer RNA, and we wrote about in the book about how her and Jennifer met each other and then they started a collaboration to try to understand the molecular mechanism of DNA targeting by CAS9. Back then, CAS9 was actually called CSN-1, so it underwent a name change around 2011-2012.

Vincent: Crosby, Stills and Nash. (laughter) That's what I think of. I don't know if you remember themselves

Sam: I know who they are. Of course. And actually, CAS9 before it was CSN-1, was CAS5 so that yogurt paper from 2007 has very clear evidence that if you inactivate the CAS5 gene, you lose adaptive immunity, and CAS5 is today's CAS9. So Martin was the one that was working with Kalinsky, a PhD student in Emmanuel's lab to purify CAS9 understand how it interacts with CRISPR RNA, tracer RNA. There is good genetic evidence at the time that CAS9 might be the DNA cleaving enzyme and in hindsight it seems obvious that it would bind CRISPR RNA, tracer RNA, and cut DNA, but at the time I mean, the pieces just weren't all there to put that together so you really needed to do these experiments to show this is actually what's happening.

So he was pursuing this in Jennifer's lab together with Kalinsky in Emmanuel's lab. They showed that indeed CAS9 uses two distinct nuclease domains to cut both strands of DNA and that this absolutely required both CRISPR RNA and the tracer RNA. And so around that time I had already set up a collaboration to look at this target search question but actually using a different CRISPR complex called Cascade, which is the one found in E. coli and it's the one that is much more prevalent across bacterial and archaeal genomes, but around the time of Martin's project coming to conclusion, I thought this might become a big tool, so I think we should think about what importance that might have in terms

of just going after systems that are not only interesting biologically but have potential biotechnology utility. Also we thought it would be cool to look at target search across two completely distinct evolutionarily unrelated protein architectures, and in fact we ended up publishing two papers and showed that they had converged on very similar modes of target search even though there is no ancestral relationship between them whatsoever.

So I was involved in the project to first show that CAS9 uses these two RNAs to make double stranded breaks, but that was going on in the lab and I just had a conversation this morning where I was recounting signing Martin's notebook pages when he first invented this idea of the single guide RNA by fusing CRISPR RNA and tracer RNA into a single transcript to turn what would have been a three component system into just two components, the CAS9 and the fused single guide RNA.

Vincent: That just makes it easier for people to target, right?

Sam: Just one less component you have to encode in a plasmid that you might be transfecting.

Vincent: So now the tracers in variant--

Sam: Exactly.

Vincent: So then the CRISPR part is gonna be whatever gene you want to target.

Sam: Exactly. And these vectors are designed so that you can easily clone oligos into them, basically to put in whatever guide sequence you want.

Vincent: Was that first tested in a bacterium?

Sam: So in the paper that Jennifer and Emmanuel published, they showed in 2012 that you could target any arbitrary site on a plasmid in a biochemical cleavage experiment. And then obviously there is a little bit of drama over which research group really invented the technology, but in at least in the public case record in 2013, there were back to back papers from Feng Zhang's lab at MIT, Broad Institute, and George Church's lab at Harvard showing that indeed if you expressed CAS9 and these guide RNAs in either human cell lines or murine cell lines, you could introduce permanent edits into genes of interest.

Vincent: I guess it's one of these leap of faith experiments, right? Because not only as you said before, the eukaryotic genome is so much bigger, but it is chromatinized, all wrapped up. I mean, that's the thing, if you think about it too much you'd never do the experiment. It's not gonna work, right.

Sam: Yeah. And it seems to work so darn well, too.

Vincent: It's amazing. Even with all that protein there. We don't understand how, right? Because there's no chromatin in bacteria, for sure.

Sam: Well, there are structures, there are histone like proteins that compact the DNA in different ways, but certainly not as well formed nucleosomes that you would think would present some obstacle to CAS9.

Vincent: So explain how we have now this complex, the CAS9 which is making a double strand break, how does that make it possible to do gene editing?

Sam: Yes, I think, I wish I could go back in time and read more about gene editing before I joined Jennifer's lab, because I would like to think I would have been working on CAS9 3 years before everyone else was. But the truth is no, I joined Jennifer's lab being interested in RNA protein biology, so I wasn't really personally on top of the decades long field to develop different ways of doing gene targeting.

Vincent: Those are nicely summarized in your book.

Sam: Yeah, I spent a lot of time after the fact reading about it but unfortunately gave it a little too late. But yeah, you know. People have been trying to develop ways of introducing breaks at particular sites in the genome for a long time, in part because of work from Rodney Rothstein which showed in the 1970s that double strand breaks are the trigger for doing homologous recombination. So it was Maria Jason who showed in the 90s that if you introduced artificial nucleases they were using homing endonuclease from yeast, if you express that in mammalian cells and introduced a specific double strand break, the rates of homologous recombination at that site went up by orders of magnitude. So ever since that discovery, there has been this idea that if you can develop the programmable nucleases, so nucleases that you can easily redesign to cleave any sequence of interest, that would be a powerful way to introduce permanent edits because of eukaryotic cells abilities to take breaks and either seal them back up, usually with some insertion deletions that can be sufficient to knock out a gene of interest, or combine with the repair template, actually introduce a precise change that the experimentalists can design.

So I think it is always important to stress that CRISPR is not the first gene editing technology and one sad part of the book is we ended up spending one paragraph or a couple paragraphs talking about TALENs, which I think would have revolutionized gene editing in a far more permanent way, if CRISPR hadn't come along and kind of taken the wind out of their sails and become an even easier technology because it uses an RNA molecule for the specificity. But the previous technologies which are called ZFNs and TALENs, those use redesigned DNA binding domains protein based DNA binding domains, which can be re engineered but they are just a pain to work with because they are not perfect. The cloning can be a real pain especially for TALENs.

If you had to invent from first principles the way that you would target different DNA sequences at will, it would be ideal to just use base pairing because nucleic acids already have a way to recognize each other through hydrogen bonding. It turns out that CRISPRs are doing exactly that, and with the work on CAS9 it became apparent that you can just have the guide RNA sequence match the DNA sequence you want to make a break in and it just works amazingly.

Vincent: So let's see. Let me give you a couple of case situations. Let's say you want to completely delete a gene from a cell line. What would you do? In terms of designing the CAS9 we will take for granted is in there. What kind of RNAs would you design?

Sam: So now you can use very easy to use paired guide RNAs where you can have guide RNAs targeting the regions flanking the gene you want to literally excise and in some proportion of cells, you will have that entire region just be lost because you will have concurrent cutting at both positions. It's important that you will have a variety of different repair outcomes, and I think some of the recent literature is highlighting the fact that you have got a variety of different repairs. In some cases even at the target site you can get large deletions that could be a major problem for therapeutic development. Some of the methods that researchers have been using to assess repairs impose a bias because if you do PCR amplicon sequencing across the site you are cleaving or introducing an edit in, you are not even going to see large deletions. So now with new methods for seeing that actually, we don't have as good of a handle on all the different types of outcomes as are possible. A number of labs have shown that you can use two guide RNAs and two cuts to excise entire regions. That's been very powerful.

Vincent: What's the maximum number of guide RNAs you can put in a cell? Can you put more than two?

Sam: Absolutely. There is a study earlier this year from Joanna Wysocka at Stanford. They developed a tool called Cargo, I forget what the acronym stands for, but they cloned tandem arrays of guide RNAs, each one downstream of a U6 promoter, and they were interested in increasing the signal to noise if you were doing chromosomal locus imaging, let's say you want to image specific regions of the genome, a single CAS9 GFP fusion, so people have done this for example using a CAS9 GFP fusion, if you use a guide RNA targeting a sequence in telomeric DNA, obviously those are highly repetitive so you get the beautiful foci labeling all the telomeres of all the chromosomes, but for non repetitive target sites it has been a limitation, how to get enough signal to noise. And they show that you can actually build these plasmids with tandem arrays of dozens of guides as a way to tile CAS9 along the region of interest. Instead of one binding event, you get a few dozen CAS9s coming along with GFP or other types of cargo.

Vincent: Amazing. If you wanted to replace an allele, what would you do?

Sam: So there you are gonna combine CAS9 with donor templates. And people are experimenting with different ways to deliver donor templates. Recombinant AAB is one of the most efficient homologous recombination. You can also put in single stranded DNAs, there are now some companies that sell very long single stranded DNAs if you wanna knock in some reporter or some epitope tag, you can also transfect double stranded DNA or plasmid based donor templates. These are often quite a bit lower. But yeah, those are gonna involve CAS9, your guide RNA, and then some type of donor template that carries the sequence you want to introduce.

Vincent: So in all cases, as you alluded to before, you have to check the cells you are working with, especially if you have therapeutic uses in mind. Make sure you got exactly what you intend and not anything else, right. And that's not easy.

Sam: Yup. It's both what exactly is at the target site and then also looking for off target edits and then you also have to worry about other ways that your editing experiment might be imposing some selective pressure on the cells. So there are a couple papers a few months ago looking at the potential risk that you might be enriching for P53 mutations because--

Vincent: Stem cells in particular, right.

Sam: Because those are gonna be better at the peak of the, you basically select for P53 null cells. And again in a therapeutic setting that is gonna be a major problem.

Vincent: I wanted to ask you about that, because P53 obviously recognizes double stranded DNA breaks which are being made by CAS9, and that, the one paper I saw said that this is why efficiency of editing in stem cells is low because of P53 and as you said you select for either null cells or with mutations and that wouldn't be good because you need most human tumors have mutations in P53.

Sam: You don't want to use a cell based therapy that has mutations in P53.

Vincent: So is that a deal breaker or are people gonna figure out what to do there?

Sam: I don't think it's a deal breaker. It's interesting because you see, it is interesting to watch the stock market because now there are three companies that are publicly traded that are pursuing CRISPR based therapies. I've never really cared that much about the stock market but now I tend to look at those every once in a while just to see how investors respond and so each time one of these studies comes out, casting some negative light around CRISPR, even though these are solvable problems, or these are things that the field is already aware of, you often see that like the date that those papers come out the stocks tank 15-20%.

I think there is a certain amount of hype surrounding CRISPR as a potential therapeutic strategy, and the truth is it is gonna be way harder than the media often suggests to turn these into drugs that will actually be effective, and that is both challenges around controlling the types of repairs you get, but I think delivery is gonna still be a massive hurdle. The same way that it has been for gene therapy. So it's easy to say that this will be a panacea for genetic disease or for cancer but it is always gonna be much more difficult than I think you're gonna write about in those first flashy news stories.

Vincent: And the stocks go down, it's a good time to buy, you know.

Sam: (laughs)

Vincent: They will always go back up. Can we talk about some things that you are interested in doing in your lab? I don't want to have you say things that you don't want to. I was recently at an interview with someone and I asked him and he said no, I can't talk about anything because it is all tied up in companies. But you mentioned earlier that you are interested in some of these other, so we haven't touched on this much but not everything is CAS9, right, CRISPR systems are divided into several types now, right? Several subtypes. So there are hundreds and maybe more. Are any of those being exploited, are you interested in doing that?

Sam: Yeah.

Vincent: To what end?

Sam: I was about to joke that I wouldn't have the problem of the company ties, but in fact there is some work I did at Caribou that, yeah, I probably, there are things I could talk about but I probably shouldn't because I am always a little paranoid about what I should or shouldn't say, but I think that the point is there are gonna be other types of CRISPR systems that will be quite useful, I think, and that have been thus far mostly untapped. And we have already seen that from the literature. Now the so called type 5 systems which encode a protein called CAS12, that turns out to be very efficient for gene editing applications. There is now a lot of excitement around a protein called CAS13, which is an RNA guided RNA targeting nuclease, so I think we still don't understand very well what types of RNAs it is targeting endogenously in native systems because we don't know that many RNA phages. So whether or not it is targeting RNA phages or maybe targeting transcripts during, oh, do you have a book on RNA phages? Oh, wow.

Vincent: I just had to show you this because this is an old, it's an old book. Published in 1975. But it is edited by Norton Zinder, who--

Sam: I've heard Norton Zinder.

Vincent: The first discoverer of RNA phages at Rockefeller and this is... there are not a lot but they are there.

Sam: You just now made me nervous all of a sudden that I forgot where I am and who I am talking to--

Vincent: It's okay, it's okay. Why would you, would there be an application for cleaving RNA instead of DNA?

Sam: So there has been a couple high profile papers that have tried to address exactly that question. One is maybe it will be a more effective way of doing knockdowns, transient knockdowns, than RNA interference. So some recent studies have kind of compared CAS13 based RNA knockdowns to RNAi and it seems to be more specific. And that's based on doing RNAseq and showing you have far fewer transcript, off target transcripts that get knocked down. It also has the advantage of being completely orthogonal, you're not relying on host machinery in the way that if you transfect SH RNAs you are relying on arginine and that can actually perturb pathways that you don't want to be touching. Whereas here you have a completely orthogonal nuclease and guide RNA.

Then in a couple of the studies they fused CAS13 to adenosine deaminases so that you can actually do base editing at the RNA level. I'm not as clear where that would be therapeutically useful but I think again as a research tool you might wanna make changes to the transcriptome that aren't permanent so they are not at the DNA level, they are more transient, and that might be a different way of studying various aspects of RNA biology. And then there is a really neat application that was actually outside of the cell where CAS13 has been harnessed as a different way of doing RNA viral diagnostics. That's a little bit complicated to explain but it relies on the fact that CAS13 is a very interesting enzyme. It both targets RNA in a sequence specific fashion but it also becomes a non specific nuclease

once it is triggered by binding to the target RNA. This is actually one of the theories behind how this immune system works, instead of cutting only the RNA that you are targeting, maybe you actually kill off the cell or induce some dormancy by becoming activated and now cleaving nonspecifically any neighboring RNAs in the pool as a way to halt gene expression more globally.

Vincent: They had a name for that.

Sam: Sherlock is one of the names for it, for the technology.

Vincent: For the cleaving any RNA that happens to be nearby, they had...

Sam: Oh, collateral damage.

Vincent: Collateral damage.

Sam: Yeah.

Vincent: There's a couple of papers in Science recently with different acronyms, yeah. We did those on TWIV, right.

Sam: Oh cool.

Vincent: Took us a long time to read them.

Sam: So I think it is another neat example of how you needed really fundamental mechanistic work to uncover how these enzymes work and I think one study that a friend of mine published in Jennifer's lab, Mitch O'Connell who is starting his lab at Rochester, he showed how this collateral damage works and that you can actually link that collateral damage to cleaving a fluorescent reporter molecule so that you can actually use targeting and read it out through fluorescence. And then Feng Zhang's lab developed that into a technology they called Sherlock where you can actually get very very high sensitivity detection of a target RNA of interest in various clinical samples like blood samples. And now they have recently turned that into a kind of, I forget what they called it, a dip strip or, similar to a pregnancy test where you can actually run your sample across some filter paper and turn on a band if your RNA is present in that sample.

Vincent: Very cool.

Sam: Now there is a company that has been founded out of Berkeley that a couple of friends of mine are involved with called Mammoth Biosciences. And they are hoping to develop this into kind of a point of care card or credit card sized diagnostic device that you might be able to use at home. Potentially with urine or saliva sample and be able to detect the presence of pathogenic RNA or DNA before you go to the doctor's office, for example.

Vincent: Could it be that diagnostics come online before gene editing, right?

Sam: Yeah, there are some people that think that maybe getting things out into consumer's hands might be faster. Clinical trials take a long time and I think the safety concerns are much higher than if you are just spitting on a piece of paper.

Vincent: Is Mammoth a play on Caribou?

Sam: No.

Vincent: (laughs)

Sam: My understanding, because I talked to one of the C level execs a year or so ago, I think it was kind of a play on George Church's work on editing elephant DNA.

Vincent: I see.

Sam: To be more like woolly mammoth DNA. I think it was roughly a play on that, but it is funny that now you have two companies that Jennifer is involved with that have random animal names. So I guess we need to found another company and call it, what would the next animal be? Something else.

Vincent: These are pretty big mammals we are talking about here. Mammoth is pretty ancient, right? Caribou is still around.

Sam: Moose?

Vincent: Moose. You could pick your favorite one. I don't know where Caribou came from, do you know?

Sam: Caribou is a portmanteau of CAS but the S is gone, so CRISPR associated ribo, for ribonuclease, and then I guess Rachel threw on a u to make it a real animal name.

Vincent: Nice, I didn't know that.

Sam: So yeah, Caribou. We have herd meetings as our weekly company wide meeting. We have what other kind of caribou themed...the in house computational software is called Tundra. (laughs) and I think there were a couple other cute little terms that played on the caribou theme. Actually we did a white elephant exchange a couple years ago and the gift I got was a book about caribou the animals, which I never read, but at some point I will.

Vincent: So after you did your PhD at Berkeley you went to Caribou. Is that right?

Sam: I wrote this book with Jennifer.

Vincent: First you wrote a book. And that took how long?

Sam: A year and a month or so.

Vincent: Let's talk a little bit about that. So I read that a couple of months ago. I guess in preparation to, well you had come by earlier this year and told me about it so I bought it and read it. It only took a couple of hours, *A Crack in Creation*. Probably took you longer to write it.

Sam: It did, yeah.

Vincent: Why did you decide to do that, I'm curious.

Sam: You know, I used to read the *New Yorker* religiously and I love science writing for non specialists. I love science nonfiction, I remember reading the *Selfish Gene* by Richard Dawkins as an undergrad and it kind of really opened my eyes both to evolutionary biology but also to kind of top quality science writing and I guess it was never a dream of mine to be a science writer myself, at least outside of the kind of scientific formal literature, but I think as CRISPR technology was beginning to crest and it became clear that this was gonna really change the way everyone thinks about doing science and doing genetic engineering, having started in the lab when CRISPR was such a small field and seeing this kind of transition of bacterial adaptive immunity into this blockbuster technology, I thought there aren't that many people that have been on both sides of that. And I think the long term ramifications are potentially huge when you start thinking about not just therapeutics and patients with disease, but some of these very controversial ideas surrounding the use of CRISPR in the human germ line.

So the fact that one could and in fact scientists have already begun experiments putting CRISPR into human embryos to make genetic changes that could actually make their way through development to a new life. And that this could really change the way that we think about genetics and childbearing and what genetic state our children will grow up with. So I think all of these ideas were swirling in my head and I remember a vacation I took with my family where I thought man, someone should write a book or a *New Yorker* story about CRISPR and so over the course of about a year I kind of joked with Jennifer here and there about writing a piece for the *New Yorker* or you should save your memories because some day you'll write it in your own memoir, and then at some point she was contacted by a book agent who is based in New York, Max Brockman, and she forwarded it to me and said hey, you are graduating at the end of the year, do you want to work on this together? And I said absolutely.

Vincent: Cool. So you stayed at Berkeley and did this?

Sam: I traveled a bit. I would say the first six months of writing were not nearly as productive as the last six months and the timing of this also happened, I wouldn't say I was burnt out but I worked pretty hard during grad school and in particular this last paper, first author paper that I published on the conformational activation of CAS9, I worked on that pretty much up until the day I left, we submitted the paper shortly before I left the lab and I think I just needed a breather so I did some traveling for a few months and kind of worked on the road, but I would say wasn't the most productive phase but I just needed to be out of the lab, be out of Berkeley, and just clear my head, and I think what better way to do that than reading books to try and put CRISPR in a bigger landscape, thinking about gene targeting, thinking about ZFNs and TALENs, thinking about the recombinant DNA revolution, thinking about the origins of DNA as a genetic material, one of my favorite books ever, probably in the scientific nonfiction literature, is the *Eighth Day of Creation*, which, yup, it's on your bookshelf.

Vincent: Purple, right there. (laughs)

Sam: I think that was such a fantastic read. Really opened my eyes to thinking about how to teach about the origins of some of these landmark discoveries and I think *A Crack in Creation* didn't end up becoming the kind of book that maybe I imagined it would be when we started the project. That was definitely affected by the timeline. One year it just wasn't enough time to do this kind of interview that I wanted to do. When I met this guy in the Netherlands who coined CRISPR, that was when I thought there might be time to go and interview all of the pioneers in the CRISPR field. I thought I could go interview all the pioneers in the gene targeting fields. A few minutes in it became clear if we are gonna make this kind of aggressive deadline that is a year from now, there is just no way that's gonna happen.

But interestingly, the first, so the title we came up with when we sent out the book proposal was *Rewriting the Code of Life*, which Michael Specter a journalist that the *New Yorker* actually used for one of his pieces on CRISPR based gene drive technology, and then the publishers came back with *The Ninth Day of Creation* as their first suggestion, which I think was a play on the *Eighth Day of Creation*. I liked the idea of it but I think a tiny percentage of readers would have understood the reference. In fact, I think I was visiting my folks when I got that email from the editor and I told my parents and the first thing my dad said was, what was the eighth day of creation?

Vincent: (laughs) Right, yeah.

Sam: So I liked where they were going with that but I think it didn't quite work. I think to be a little bit provocative they wanted to keep this creation term in the title because you've got creation and then the subtitle is *The Unthinkable Power to Control Evolution* so it really touched some buttons, especially for readers that might think about creation vs. evolution and be putting kind of this power to rewrite DNA into the context of who is really the one that should be making decisions about the human genome.

Vincent: It's interesting because it is written in the first person of Jennifer, who is coauthored. Is that something you decided early on because she had been given the book proposal?

Sam: So the whole story behind writing the book, we had this interest from the agent, we had a phone call with him, then I had to go write my thesis and go through the process of actually graduating with my PhD at the same time as writing the book proposal. And at the time we thought we kind of were already in the front door and it was a formality and they would maybe help us edit the proposal, but we put something together that I think I was proud of, but didn't approach it thinking this was our chance to really seal the deal, and so the agent actually passed on that first book proposal. That first one was written from the perspective of we, and I think one of the weaknesses was it wasn't really clear who 'we' was, because in some cases I used we to mean Jennifer and I. In some cases it was we, society, or we, scientists, and I haven't gone back and reread that first draft of the proposal but I could imagine it was hard to connect with the writer because you didn't know who are you actually hearing from?

So when we revised the proposal, I think I also, it was always clear that Jennifer is the celebrity scientist here. She's the one that the publishers, the agents, want to hear from and who the readers

were gonna want to hear from because she is the one widely seen as the coinventor of this technology. So it was important to me to not get involved in a project where I'm gonna be a ghostwriter who is not even on the front cover, but it was clear from the beginning of the project that I would be the one helping to do a lot of the legwork and authoring the book, but this is Jennifer's, her telling her story, and I think by the second book proposal which was much tighter and I think had a more clear target audience and a much more clear voice. We decided it really makes more sense for me to write this in Jennifer's voice and as we talked about the prologue, I mean, I'd say 99 percent of the ideas and the opinions of the book are really shared between the two of us because we think very much alike on a lot of the different issues that we get into. But there are also things that are about Jennifer's life or about her getting in touch with Emmanuel and starting this project and I wasn't on that first CAS9 paper so it really made more sense to tell the story from her perspective and explain in the prologue why we went with that decision.

Vincent: I think that the idea of interviewing people is great. When you get tenured, maybe you should do that.

Sam: I really had lofty aims to do, to write a book in the style of Horace Judson of the Eighth Day of Creation, and the cool thing is I know virtually all the people in the CRISPR field that were involved since the beginning. One thing the book I think didn't succeed in doing is celebrating everyone's contributions in the way that I think you read the Eighth Day of Creation and a) you feel like you're there and you can really reset your expectations for what was DNA. What did DNA mean to people in the 1930s? And I think it would be cool to try to put people in the mindset of what do people think about CRISPRs in 2005 when you didn't take it for granted that of course that's what they are doing, of course CAS9 makes double stranded breaks and target DNA. And I think there is a lot of fun little vignettes involving all the different researchers. It would be fun to write that book. At the moment I am trying, I'm more focused on doing science and getting tenure, but maybe, someone will write that book. Michael Specter is writing a book and there will be other books out there.

Vincent: Could be. But you know, you go to meetings you can bring a little recorder and just take a half hour and sit down with someone and just archive all this stuff and then at some point in the future you'll have it. Doesn't have to be done all at once. The interviewing can be done whenever you're at a meeting and there's somebody new there. You take a half hour and you could spare that. You don't even have to have a plan, you just get everybody. Talk to everybody and then later, you could put it all together. One of the really good science books is by a guy here, Siddhartha Mukherjee. Right, the Emperor of All Maladies. He did that when he was a resident.

Sam: Yeah, I don't understand that. I actually met him my first or second week here. It was funny, I was having lunch with Uttiya Basu and we walked by this Thai place on 168th street because he said it's always too busy and he kind of commented, oh yeah, and I always see Sidd Mukherjee there. So we ended up eating lunch somewhere else. The next day, I'm having lunch with someone else and we did go to that Thai spot and ten minutes later, Sidd Mukherjee walks through the door. And he had actually just done an event in Berkeley with Jennifer, I think a week or two before that, so I had never met him. I'm still kind of in awe of the guy because of his research, his clinical work, and of course his two books now. One of which won the Pulitzer prize, I think.

Vincent: The Emperor won a Pulitzer, yeah.

Sam: So I went up and introduced myself and he remembered my name at least and then we had a meeting in his office a couple weeks later. But yeah, I don't know how you could do that while you are a resident, that kind of blows my mind. I had a hard enough time writing a book when I was doing it full time.

Vincent: Let me just ask you one more series of questions which have to do with your subsequent career. So you mentioned that you needed a break and the book was one way to do that. What were you thinking long term? You went to Caribou but were you thinking about that when you were writing the book or when did that come into play?

Sam: I don't know if it will make me look bad to not talk about having had a vision for the beginning, but the truth is in grad school, pretty much up through the end, I was still dead set on becoming a PI in academia and mentoring students. That has been my passion since starting research in Ruben's lab as an undergrad. Then the book felt like okay, it's a slight risk because that's not a typical thing on the CV of professors, but I thought I'll never have an opportunity to write a book like this and get paid to do it. Other people write books but I had the benefit to do this with Jennifer and she was someone that publishers were more than happy to pay to author a book like this. So I kinda thought it's a risk for my career but that opportunity may never present itself again. And then the plan had been start a post doc right after I'm done or maybe apply to some of these faculty fellows programs, but truthfully finishing a book was a major challenge for me and I think I knew six months out that there is just no way that I can mentally balance finishing this manuscript and also answering this question of what I am gonna do next and who I am gonna do it with and where I am gonna move to.

So in the end I got to September of 2016 and I just didn't really have a plan. I just had the manuscript finished and then I thought I've gotta figure out what is next now and I didn't do an advance but that's fine. And so Caribou was a company that I had already talked with before the book project about potentially working at. I knew the CEO really well because she worked in Jennifer's lab. I already had a bunch of meetings over there. So it was something that had been in the back of my mind as a potential position. I wouldn't say that I was passionate to explore a life in industry. But I was back in Berkeley at the time, I thought I'm curious about industry, I would be able to stay in Berkeley, I would be able to get back into the lab as fast as possible, which is something that at that point I was done with the whole living the life of the author like waking up and sitting at a computer all day. I wanted to be back at the bench.

And yeah, I just thought this is a good next step. It's not necessarily my passion in life to be at a company but let me start and figure out the rest afterwards. So then I started there and then the month before or a few weeks before I started, I thought I had a bunch of phone calls with advisers and friends and I decided I'll apply for a faculty position that same Fall. Having kind of skipped a traditional post doc, but there might be little to lose to just put some applications out there. I had the fortune to kind of get a lot of work done when I was in Jennifer's lab and riding the CRISPR gravy train. I mean it certainly helped with that. So I thought let me see if anything bites.

So I sent out, I started putting together a research proposal, sent out a couple of those in applications, and then was really thrilled when I got the offer at Columbia, because not only is this a great place for me academically and intellectually, but having spent six months of my PhD working on this campus

and having been here as an undergrad, I really feel like I have a community here that preceded my arrival as a professor and it is really nice to work somewhere where you have connections, you know some people, I knew that I loved New York. So it just actually felt like the perfect place for me to come back to.

Vincent: So at Caribou, were you working independently, more or less?

Sam: I was a group leader of the technology development team. So when I joined it was maybe 40-45 people. So I had some people on my team, actually one of the interns that I worked really closely with is now in my lab as a lab manager and she'll be starting a PhD program next month. So what I really liked about Caribou is that it was still some place that I could work closely with people, mentor people, so I had a team of a couple, maybe 4-5 people. I had kind of the main project that I was leading but I was involved in a couple other projects. I have to say, a great work experience, I loved the people there, I really miss working there, and it was a neat industry experience because in many ways it felt not so different than doing academic research. I think at the time they weren't really in a position where they had to churn out some product. So there was still a lot of fun exploratory work going on and the work that I did there is not so different than the kind of work that I might have done in Jennifer's lab or that I will be doing in my lab here.

Vincent: It's good for people to hear your story, because they can learn you don't have to do the traditional post doc to have a career. You can do other things. I always tell people, try and work out of the box, you know. Do something different.

Sam: I think looking back I'll be, I feel happy that I didn't stick to this very linear path that I think in academia would feel like we need to be on. It's gotta be PhD, it's gotta be post doc. I mean, actually, kind of there are many post docs that I would have liked to do as well. One thing I wanna keep doing is keep on learning so I feel like while I'm starting the lab I still want to figure out a way to do post doctoral research by just teaming up with different people. It's always fun to keep on learning new things.

Vincent: Well now you have what is a dream of many scientists. You have your own lab and academic place, good place, welcome by the way. You can spend the rest of your life doing multiple post docs and living vicariously through your post docs.

Sam: That's right.

Vincent: Alright, that is TWIM 184. You can find it at Apple Podcasts, microbe.tv/twim. If you listen on a phone or tablet which is what most people do, you use an app, you can search for TWIM. Please subscribe, it helps us if you subscribe, it helps our numbers. If you really like what we do consider supporting us financially. Go to microbe.tv/contribute for the different ways you can do that including a Patreon account. Of course if you have questions and comments please send them to TWIM@microbe.tv. Sam Sternberg is right here at Columbia University Irving Medical Center. He is on Twitter, [shsternberg](https://twitter.com/shsternberg), we'll put links to his websites in the show notes. Thank you so much for talking with me today, I appreciate.

Sam: It's a lot of fun, Vincent.

Vincent: I'm Vincent Racaniello, you can find me at virology.ws, thanks to ASM for their support of TWIM and Ray Ortega for technical help. Music is by Ronald Jenkees. Thanks for listening everyone, see you next time on This Week in Microbiology.

(music)

Content on This Week in Microbiology (microbe.tv/twim) is licensed under a Creative Commons Attribution 3.0 License.

Transcribed by Sarah Morgan.