Transcript

Hello, and welcome to *Doing Science Differently*, a podcast that explores issues in the culture and practice of research. We interview experts working on making the world of science a better place and learn how their pragmatic approaches can change practice in the lab or clinic.

Today, Eric Danner will be talking to Tim Errington, the Head of Research at the Center for Open Science and manager of the Reproducibility Project: Cancer Biology. They'll be talking about replicability and challenges specific to life science experiments, and about what we can do differently.

Eric
So, can you start by introducing yourself?

Tim
Sure, absolutely. So, I'm Tim Errington. I'm the director of research at the Center for Open Science. We're a nonprofit based in Charlottesville, Virginia that has a mission to increase openness, rigor, and reproducibility of research, and that's agnostic to any disciplines of research. We've been around for about eight years, since about as long as I've been at the organization. I joined after I finished my PhD at the University of Virginia in the Microbiology, Immunology, Cancer Biology Department. The Center focuses on three main activities: meta research, which is what I lead and the research team, infrastructure, so developing open tools and software that enable more rigorous reproducible practices such as the Open Science Framework, and policy working with communities and stakeholders, such as journals and funders to kind of align those incentives so they match those ideals of science that we're aiming for.

Eric
Super cool. So, this work that you just did. You worked on it for about eight years, nine years? A while. So even though I think your original plan was more on the five-year timeframe, this was still a huge undertaking. And so, I'm curious what got you so invested in this question of reproducibility and cancer biology?

Tim
That's a great question. Actually, it gets to why I joined the Center to begin with, and it's a little bit of a happenstance or luck that goes into that. So, as I was just saying, I finished my PhD at the University of Virginia, which is in Charlottesville, where the Center is located as well. And in 2013, right when the Center started and when the cancer biology received funding, the idea received funding from the Lauren & John Arnold Foundation, which is now known as Arnold ventures, I caught wind of it. I think I forgot where I saw it. I think I might have seen it like in a *Nature* ad or something like that. There's something saying, 'oh, there's this new project that just got funding', right? It's really an announcement that this project got funding, that this company 'Center for Open Science', this non-profit, is leading it. What is this thing? It's in Charlottesville? I live in Charlottesville. Like what is this brand new organization? I google it, right? I'm like, wow. So, they just, they just received a whole bunch of money to do things in reproducibility. And they're going to now look in preclinical cancer and I was like, this is excellent. Like, this is exactly what I think we need to do. Because again, the Bayer/Amgen studies had just come out and definitely caught a lot of wind in our department. And so, I remember coming in, talking to Brian Nosek, our CEO. And the office, there's nobody there. I

think when I walked in one other person, Brian, was hanging out in an office space. And I remember sitting down he's like – oh, what do you want to talk about, right? And pretty sure I talked for the whole hour, nonstop. Just about my own interest in it. I was like, I definitely remember those Bayer/Amgen papers, I was like, I've struggled with it myself in the lab, like, you know, I first started my PhD at UVA, I was at Berkeley beforehand. And I went to UVA, I spent the first six months trying to reproduce somebody else's experiment and the paper that just got published. And that person was in the lab, I had access to like all their notes, to the person and could talk to the person. Couldn't figure it out. And so, I spent months trying to be like – how did you do this? How did you do this experiment? And we couldn't figure it out. And, again, eventually we figured it out because I had to do basically detective work. And it was a methodological detail that got lost along the way. It got recorded early, and it just stopped being recorded. So, I started testing all these conditions, and then like, finally figured it out. I was like, this is horrible. This is not how we should operate in science, it's a lot of time I just spent just trying to like repeat one thing. And I was like, and I had the person, like the methods of the paper were horrible, because they're like – oh, go look at this other paper. So, I remember talking about that. And I just left, talking with Brian saying, I think this is exactly the way it should be done, open, let everybody scrutinize the data. Like there's good things we're doing, but there's definitely things that we can improve on. The only way you can do that is to do it in the open. I think that's wonderful. And I was like – I want to help no matter what happens. And so then, the next day, he called me up and says – do you want to work at the Center and lead that project? And I was like – absolutely, I think that'd be wonderful. So, my investment came from my own background, my own PhD experience, my own research experience, struggling with it. And again there's good things, but it really is to me, what always drove me nuts was just not understanding it, not understanding what was originally done. The methods just being so sparse and having it be such a pain to figure out – well, what should I do when I'm designing experiments? Like, where do I start? What reagents do I use? Nobody really, you know, records all the controls for, you know, a knock-down experiment, for example. So that used to just frustrate me, just the inefficiency of it. So, I brought that to the organization and to the project when I started. So, I think to me, it's been a little bit personal just because I know exactly how it felt. And I think the project took so long, probably because of all those issues that I experienced, they just kind of came out in full force. It took a lot longer than we thought. Because of the challenges that we described, you know, part of it is also remember that it's really hard to always know, in any research. This meta research is no different than the bench research that we do as scientists every day, it's hard to predict how long an experiment takes. So, I think this is just one big experiment, you know, we made a guess in the beginning, received some funding, thankfully, because that's really hard to get funding for these types of studies. And we got it wrong. And we just can't predict it all. And so, to me, what ended up being more important was – how do we finish it to make sure that we learned the most, right? Let's make sure that the quality of what we're doing is the most important, not the timeline. And I think that's true for any scientist, right? Our goal shouldn't be, 'I must finish this and get a paper by a certain period of time'. It can be, but I think that's the wrong goal. The goal should be, the resources and the time I have to invest in this – how do I make sure I get it as far as I can? So that way, I can hand it off. And to me, I felt like that was the, and the organization thankfully supported us in saying that, that's the point. The point is to finish this to make sure it's the best it can be with the resources we have.

Eric
After these nine years, you actually just came out with this series of papers, to at least bookend this chapter a little bit. And to me, it seemed like there was two major takeaways. The ability to actually reproduce, and try and carry out the experiment. And then once you could carry out the experiment, was it reproducible? Can you start by telling about the reproducibility of the experiments you were actually able to carry out?

Tim

Right, yeah. I think you're right to focus there. So, if we just look at what we had outcomes for, right? What we always latched into, which is – well, you found X, did I find X? What we found, and this is a really important question, by the way, because it's not – we want it to be binary. Yes, it's reproducible. No, it's not. We know it's not, it's incredibly complex. And it really comes down to what one wants to find, to what matters, is another way to say that. And that's going to vary depending on the specific experiment and one's interests. What do I want out of it as an individual, right? Do I care that there's an effect? Or do I care how big it is, for instance? So, we do have to kind of remember that we have to gloss over that a little bit when we look at this in aggregate. And, to me, that's actually a fun and exciting question. So we tried to do that, I think, well, hopefully well enough in the paper, because there's many ways to look at it. And one of the ways I think that's the best to ask this kind of really difficult question that I get asked a lot – how many are reproducible or not? – is looking at all the different types of methods that one can think of, at least looking straight at the data from a quantitative standpoint. And so, we looked at five methods that kind of work really well for this. And then we said – who cares which one, out of those five methods, did the majority of those, you know, how many experiments did the majority of those kind of get similar results? And so, this is where we come up with this 46% number, right? This 46% that we were able to replicate. We're really saying, that's not really the way that, it's a nice way to distill a lot of complexity down into something. But I think that works really well because it's saying – regardless of what you find important, we're saying less than half kind of met a fairly loose bar. And of course, you're welcome to go up and down the ladder, you can be as critical or as not critical as you want when you do that. So that's one thing that I find, which is, it's really hard, but you do find that, thinking that you get exactly the same, very few map that way. The majority of them are a lot more complex.

But the thing that really strikes out to me in the entire project, especially considering the domain area, this preclinical cancer area, right, which is, which is a feeder into the clinic, right, which is our hope, as scientists, right to saying – we really hope our finding will one day map its way out and actually make an impact on human life. So, to me, the magnitude of our effects is what really matters, right? When we say – I found drug X, you know, shrinks tumor by Y – the size of Y really matters to me. Because that's really the impact that we're trying to say. It's not just that I see something, it's that I see something that is worth taking notice for. And that to me is probably the biggest finding that we had, which is that 85% of those effect sizes are smaller, 92% of the effects are smaller than the original. 92! I mean, that's, that's really staggering to me that was more than I thought, you know. I assumed that it, you know, we always hear about, like, you know – there is regression to the mean, that always does shrink, and that we have very small sample sizes by and large, so lots of noise. But almost everything was smaller. And that just I think, is probably the bigger thing to really take home for me, which is, when we did these experiments, when we did find outcomes, yeah, we could see things in the same direction, we could see effects, but most of them were not the large effects that we got excited for in the first place. And I think that's something to really think about as we take these findings and move on. Especially since like I said, the goal for many people, at least for me, I think the most important goals, the hope that this translates one day into the clinic and make an impact.

Eric
And when it doesn't have as dramatic of an effect – how much smaller was this effect?-

Tim
Yeah, yeah, so if I'm hearing you right, on average, we saw that they were 85% less. That means, you know, if we saw something that was 100, a difference between, you know, control and drug showed a difference of 100. We saw a difference of 15, right? So, we saw this pretty substantial decrease in the size we had. Something to keep in mind there, and it really is nuanced, like one question that comes up, I think that's related to what you're getting at, which is – what causes that? What's the reason? Why are some of these smaller and some of these are not so small? We don't know, to be quite honest. We set the whole project up, hoping to figure that out. There were a bunch of different, you know, moderating effects – maybe it's got

to do with, you know, whether the authors helped us, right? Maybe it's got to do with how well the protocols were described? Maybe it has to do, if we use the original materials versus recreated them, right? Like maybe their mouse model, it works, but when we got a new mouse model, it stopped working. So, we tried looking at a number of these things that we thought were important, and nothing really popped out. It doesn't mean that they're not important, it just means that we didn't, we can't explain the results that way. The most likely thing that sticks out, though, is the size of the original effect. So, if the original effect, I mean, this is kind of counterintuitive, right, if the original effect is large, there's a good chance that we probably found something that was more substantial. And if the original effect is small, we had a harder time finding it. And that's kind of known, that's like the whole power debate, right, which is like, small sample sizes are notoriously difficult, in terms of getting good precision of estimates.

But I wish I knew, because I think that's actually something really important, right? Which is like – what is it that makes the findings kind of be more true? And more true in terms of the effect size between the original replications. And, sadly, we don't know the answer. I think to me, it's probably everything, right? There's so many factors at play. It's like research, right? Like we can't, we'd have to do a lot of experimentation, I think, to pin it down. And that's both, you know, unsatisfying in the sense that, like – shouldn't this paper just be the one that tells us the answer? I think it's satisfying in the sense that it tells us, yeah, everything that you've been hearing from a lot of people about making sure you're randomized, making sure you're blind, making sure that you, you know, have good sample size, actually good chances, and, like, your, you know, your controls are good, good chances are all those things matter. And we know that. I think the more important thing is, to me at least, is – let's also make sure that we're reporting everything. The thing that we can't say here is also – did the original papers – and again, this isn't trying to suggest anything, but it is something that's known, which is – are the original papers showing us everything**,** right? Like, – are they trimming the data, are they only presenting the experiment that quote unquote 'works'? And what we did is present everything. And now what we're getting is the truth, right? We're getting a closer thing of like – well, actually, that's probably what they saw in the original, they just didn't want to show us that. Because it doesn't get it in *Nature*, or *Cell*, or *Science*, unless it is a huge effect. So that's something to keep in mind. We know about publication bias and selective reporting. And I think that is potentially something that's at play here.

Eric
Do you think that, now when you're looking at papers, do you actually just think that – oh, it's probably 80% less and I only give it 50% confidence? Or do you think, if I tried to do it, I would get a smaller effect? But that, like, because, of course, one of them is a fundamental truth that science is trying to get to, and then one of them is a technical replicate issue as it gets passed from hands to hands. And I guess, it is not exactly like, technically able to parse these two, but just as a feeling, like, how are you viewing this? Do you just approach it all now, you know, like – ooah, this is, the system is uncomfortably problematic?

Tim
Yeah, that is such a good, that is a good question. Yeah, no. I've not thought about this. I think about it slightly differently, I'll tell you that. But I've not thought of it this way. This is brilliant. Um, I think if on the spot, I would say I'm skeptical. Yeah, absolutely. I think now this project opened my eyes a little bit wider than they already were. And I thought they were pretty open. And yeah, I think when I see things that are in the literature, I, probably now very much I'm like – that's awesome but It's probably not that big. It's probably much smaller.

Eric
And on top of that, if I tried to do it…

Tim
Yeah, exactly. If I tried to do it. Yeah. Like, I wonder what issues I'm gonna have? Yeah, no, it is easy to become skeptical. But actually, there's an anecdote there. So early in the project, we had help parsing methods from papers. We kind of, we had a bunch of people that helped us. On the registered reports that were published, everything's been published and open, there's always this one author in there. I know who they are. That was a volunteer that helped us with the methods and we gave them authorship because they basically, they tried to, they teed it up for us across all the papers, which is: go to the original papers, follow our process, and looking at certain experiments, and you try to help us figure out what's known and unknown from the paper. Which is, by the way, a really important exercise, you should try it. It's a, it really opens your eyes to how well or not well methods are reported.

Eric
Wait, sorry, can you say it again, what did they do?

Tim
Yeah, so they would, we would pick out experiments from the paper. So, we said: here's these papers. Here are five, six experiments among the dozen that are in the paper. So, we told you, we tell you by figure, that's the easiest way to do it. We're like, we want to replicate Figure 2A to D and Figure 3F. So those are what we picked out. And then we asked these volunteers to come and help us identify the methods, to say – what did they actually do? Like, we see they reported some findings. They used some cell lines or some models or mice models. Can you tell us what they did? Go to the Material/Methods section, go to the Figure legends, read the paper, and you tell us and you reconstruct a protocol for us based on that. And that's where we'd start with a lot of this, we'd start with the paper, identifying these experiments and trying to figure out what was done. And as we reported in our final summary paper, like, it never worked out. We always, there were always gaps. There was always something unknown, and it varied in degree. And I think that in itself is a really good exercise anyone should do, which is: go to your favorite paper, and try to reconstruct the protocol. Try to see if you can figure out exactly what they did. And what I think you'll find is that it is widely variable. And I think we all know that. And there are some that are just shockingly missing details where you had no idea really what they did. I think there's at least one experiment, I remember where I found nothing, nothing in the methods that described it. I think the only thing I found was maybe the figure itself is like a sub figure. I was like, I don't even know what mouse model you used. Like, I just see a result, I see a figure that shows me something important, but I've no idea what you did here. I cannot figure it out for the life of me because there's no details in the paper. And that's the extreme in the wrong direction.

But in that process, when we had all these volunteers helping us, there was just one voluntary member who had a really hard time with this. Had a tough paper, and tough experiments not well described, and just really kind of struggled with this. And I remember this email, I still have it in my email account, which is – I can't do this anymore, Tim. Like, I've gotten to the point where I don't read methods anymore. I don't believe they're going to help me. I look at papers for inspiration, not for truth anymore. And I was like – oh my gosh, it's gotten that bad, hasn't it? Oh my gosh, like we look at our papers, not to actually tell us something but just to inspire us to figure out the next thing to do. They're like, the methods don't matter, right? Like none of it matters, it's almost like it doesn't matter anymore. And isn't that sad? It's like, it's really like, I was like – my goodness, this is why we're doing this project.

Eric

Yeah. You had a quote in your paper that said – are we looking for, that we trust the data or that we trust the person? Or are we creating, like, a web of trust of the people? Or can we actually say this data is standalone trustworthy?

Tim
That's right.

Eric
And, and I think that's one level more. It's like a third branch, which is – I don't trust, I'm skeptical of the people and the thing, but I'm just going to be vaguely inspired by the abstract idea.

Tim
Yeah, right. Yeah. Science. Science is a show me enterprise not to trust me, right? And it really is. Like, you're meant, our job as scientists is to show each other in as much detail that we reasonably can do, what we're doing. It's not trust me – Yeah, the methods aren't described, but trust me, it's good. Like, that's, it's not. Science isn't built that way. It's built off transparency.

Eric
Yeah. So this kind of, this conversation about the struggles that you immediately, I mean, when you decide to reproduce something, the first thing you do is go and read the method, and try, and create a little protocol. And this kind of is a nice segway into the second concept, not of – if we do the experiment, is it 85% less, when we actually measure it, or even 50% of them don't have the effect statistical effect at all, but rather, the journey to even be able to run that reproducibility study, or to replicate it or I mean, like a normal scientist would do, to see something really cool and try and implement it in their lab in their own hands. And this was maybe the other half of this whole project, is discussing the journey to try and do this, and the complexity involved there. So, to summarize, you originally selected 53 papers, managed to pre-register 29 of them, and then you finished 17 of them. So, can you kind of unpack this? Because I think there's a lot there with how that played out.

Tim
Yeah, there is a lot there. And we tried our best to present it. And use the different methods to do that. A lot of what we put in there are case studies to try to illustrate some of the points. So, I like that *eLife* was able to follow through and figure out a way to present that in the paper. So, there's two, there's two big factors that are at play here and they're connected to each other. And one's not terribly exciting, but important to remember. The other one is probably the one that is going to catch the most attention. So, the two things that are at play. One is, one is mundane, which is we're doing research. And things don't always work out the way we have. And we have finite resources, time, personnel, money. And our job is to juggle that. So, at this big project, we started out with, you know, all these papers and experiments. We said – alright we think we can do this. Having no idea if we could, right? Nobody sat there and documented, and nobody told us how much to do this, especially going forward. We have to remember those Bayer/Amgen studies, like those were not like somebody decided to go forward, they just look behind them. They're like – we've been doing this for 10 years, like what's been going on over the last 10 years? They look behind them in a retrospective study. That's what those papers are. So, we were blind, trying to say – how much does an experiment cost? How much? Do we even know that? And we had to juggle that. So, when we ran into struggles, it was just this decision matrix that was kind of evolving in front of us, which is – oh my gosh, this isn't playing out the way we thought, this is taking more time.

And this is the other challenge, right? Which is, it didn't play out. And some of the reasons we had to make these tough decisions of how far to pursue an experiment or not, was dictated by the challenges that we ran into. Do we know all the methods? How much time do we have to recreate methods? The worst is the reagents, right? We had to remake a lot of reagents. There was one paper that we didn't pursue at all. And we didn't pursue it all because the experiments required a transgenic mouse model that the original authors made, right? Which is a huge work. But they didn't share. They didn't deposit it anywhere, and they wouldn't share it with us. And so, we're kind of like, sure, like – well, what are we supposed to do, like we make this? This took you years to make, right? Like, there's no way we're gonna sit there to just replicate your experiment and remake a huge mouse model. So, we just stopped it, flat out. Like some of those stopped for very horrible reasons like that. But other ones stopped, because we had to sit there and remake, you know, cell lines or remake plasmids. And it just, it adds up, if you keep doing it over and over, and over again. It adds up to the point where we're like – well shoot, now we're spending a lot more on some experiments, ‘cause you want to do them as best we can. And now the other ones are just sitting on a shelf. And so now we can't pursue those because we do have finite resources. So we had to struggle, this was a balancing act between the two the whole time, which is we had a desire to do a lot with resources and an attempt to do it. And those challenges that we just kept running into over and over again, stopped us, either just flat out.

Eric
You're kind of implying that somehow it feels like you're pointing at yourself and saying we mismanaged our expectations and time. But another way to read this would be, we overestimated how well the method sections were. Actually we overestimated how people would deposit their plasmids, make their cell lines available, how much they would actually share and how transparent they are. So, I feel like maybe years of doing this politically, you've been able to turn this, and be like – oh, we made some mistakes.

Tim
That's a very good point, good point.

Eric
If, do you think that the 50, now that you look back, if there had been more clear methods and most of the reagents had been shared more quickly, would 50 have still been realistic? Or is it just, 50 is not realistic given how much people actually share?

Tim
Yeah, so first, way to catch me on that. It's both by the way. And it's not us mismanaging, it's just us the way we manage. So, thank you for calling that out. I probably, I have gotten quite political at this. It's interesting you use that word. Your questions about, that's another good one, right? Could we do it? So, I'll answer that, and then I'll give you a different answer as well. I'll answer that, because I think yes and no. I think, yes, we probably would have accomplished a lot more, if we didn't have to have all these struggles. It would have been a lot faster. We would have been able to use our resources a lot more quickly. But go back to the anecdote with me in my lab, right? If I didn't have to spend six months reconstructing one methodology and one experiment, would I've gotten more done? Absolutely. I would have used those resources, my time being probably the most valuable, in my opinion, to do something else. I would have advanced a lot farther. Same things for us. If things were described better, we didn't have to recreate reagents, we didn't have to spend time trying to figure things out, we would have used our resources elsewhere. And we would have advanced more. In our case, you can measure it. It'd be a number, like you said, closer to 50 than it was to where we landed. So, I have no doubt about that. Whether we'd still achieve it, is a separate question. And that's

because it's about funding at the same time, right? We don't have infinite resources. Nobody does. But we don't, it's not like we're, this is an area of research that we were going to continue not forever. This was a project, we had finite resources for it. And our goal was to see how far we could get with the resources. It was an investment question that a private group gave us, no federal funder gave this, right? It's not like NIH or NSF gave us money to do this. It was a private group that gave us funds to say, yeah, go see this and do it in the open, see what you find. So, I do think we'd do a lot more, a lot more and a lot more efficiently. We'd have been a lot faster. That's probably the bigger thing. I don't think it would have taken us this long. But I don't know if we'd finish as many.

Eric
Do you think that, because there is some parts of your papers that talk about how things have progressed, these papers were from 2010 to 2012. And there's a lot of self-criticism about this in the last 10 years, and journals have altered it, and funders have altered what is required in terms of method and sequencing information. But do you also throw out this idea that roughly getting the core experiments is about $50,000 per paper, I think. So, what do you think that this is? If for others or for agencies that are looking at this, do you think this is a realistic number? Or does that not capture? Is there, is there something that that's not capturing?

Tim
Yeah, yeah. Okay. So, if you go at that number and stuff like, that's a really good question, we should be very mindful of. If agencies are looking at this don't over-, don't over-anchor on that number. The numbers meant more as an illustration of what we could measure. So, we were keeping as best notes as we could throughout the process. And budget was clearly one thing we kept notes of. But what's missing in that number are two huge things. So, we actually, we talked about it, if you read that whole appendix, one is missing numbers. We had people, we had reagent organizations, Sigma and all the rest that donated reagents to us. So, we never captured those costs. So, there's, you know, hundreds of thousands of dollars throughout the project, if not more, given through animal models and cell lines and key reagents that some of these companies gave us. So technically, that 53 goes up because of that size. How much, I can't tell you, but I know that's more. The second thing is, I, we don't count like my time, right? So, like, I'm not included in that. And so, myself and everyone else that's managing it, like we still play a role. So, all the admin, if you want to think of me as an admin person, I guess, I guess so. I don't really think of myself as admin, but I'm missing. And so that's a hidden cost. So, this is more of a straight cost to the providers on science exchange platform that we could measure, because that was a transaction that was very easy for us. It's like, it's like our receipt. It was easy for us to keep tabs on that receipt and keep an eye on and so we did it. We had to manage the project throughout it. And then we thought it'd be a good use case to say – here we'll show you what we're talking about, how it started, and then just kept going up. And it kept going up because everything else we were talking about. But yeah, I wouldn't, I wouldn't over-anchor on it. But on the same note like, again, does anybody else do this? If we think about the investment of a traditional, like R01 Grant and NIH in the US, right? Like, these are million dollar grants. These papers are built off of years of work. I definitely got told that a lot in the beginning. We spent millions of dollars doing this paper and you guys are trying to do it for tens of thousands of dollars. And my response to that would be well, one, we're not trying to do your paper and two, didn't you do all the work to get us to the point where it's a handoff, right? Like we're not trying to start and recreate and reoptimize conditions. We're trying to take your conditions and run with it. So my expectation is, investing, say, 3 million in the paper, if that's what we're going to use as a benchmark, I don't need to do another 3 million to replicate it, like that's inefficient in my mind. It should just be tens of thousands of dollars. We're just trying to say – do it again. That experiment's done that way, can we do it again? What we find is you can't because we're missing so many details. We didn't understand what they did. But I do think it's a good question to raise it because more agencies probably should start looking at this. I know DARPA in the US does this in some of their programs, where they fund these

like independent groups at about 10% of the cost, to basically do it alongside them and say – we're gonna replicate your results and you can't continue until we get the same finding.

Eric
So one of the interesting things that you did is, setting it up ahead of time as a registered report and getting it pre-accepted before you guys, or somewhat pre-accepted, before you actually go out and do the experiments. Can you speak a little bit about the pre-, like how that works. But also how the idea of pre-registration is, once you have it and you've done it, to a degree it's accepted. But it sounds like there's quite, there was still some challenges for you at that side. Could you speak about the pre-registration and any, like, examples of that process, and what that actually looked like for you?

Tim
Yeah, yeah. And this is a great format, *eLife* was a great partner for it. And again, it's nice to see that publications standalone by itself, because I think that provided, and I'll get to that in a second, provided a lot more additional oversight. So, this concept of using registered reports, of getting the peer review beforehand, it was essentially a complement and a better complement, I think, than the way that we were doing it informally. And it gives kind of a couple things there. So, we were always reaching out to the authors before we wrote up a registered report. But it allows those authors to engage in the peer review system the way it's traditionally done**.** And I think that having an independent group like *eLife* manage that helps a lot. It helps engage, it helps one maybe engage the original authors who might be hesitant to engage with us. Again, we don't know how many accepted. We know that they always extend an invitation, but I think with the exception of maybe one person, nobody ever disclosed their name in the peer review. But it gives them a chance to come back and say – let me give you more details. It lets other people weigh in, which I think is very valuable. And that's the really important challenge. What should we replicate? What experiments should we focus on? And are those described sufficiently? Are we doing a good job? Did we miss some detail? So, the nice thing about that is, it provides a lot more oversight into our own assumptions. It kind of holds us as researchers in check, right? Which is, we're reading the paper, we see lots of gaps, we can't figure out all of them to fill in, we're trying our best. Please tell us how to make it the best before we invest any research into it. Because again, our job is to do the best job at what we're doing, not to stumble and fall, and then get upset along the way. So, I really like that. So essentially, what it builds in is this pre-commitment. It's saying we, as the researchers, you, as the reviewers, the editors, everyone involved is pre-committing that this is good. We did it, we think we've designed this with enough detail, that there's no reason to think beforehand, we're not going to be successful. And that's really nice to have that.

And then, what I was gonna say is, *eLife* is publishing it. And being open access also means that the community can look at it. And I know there's at least a couple times when we've had some people call out and check things for us. There's one paper, where we, thankfully it was just a mistake in the paper, so we corrected it. But there's a sequence that was wrong. One of the peptide sequences was off. And somebody in the community was like – that's wrong. You guys have a mistake. And we're like – oh, my gosh, you're right. We missed it. Everybody missed it. And then we were like – oh, it's just thankfully an error. What we were actually doing, because we'd begun the experiment, was right. But thankfully, this person in the community caught us. And we would have just put the brakes on the whole thing and fixed it. And then, I think there's another one where we had that, where some of the original authors came out, said something afterwards. We were just getting ready to start the experiments. Paper got corrected. And we just quickly adjusted things to fix the errors that they found. Right before we started, and to me, I think that's the whole point, right? Isn't that the point of science, right? Share it, get the feedback, hopefully at the right time. And if not, then at least acknowledge what's missing.

So that process worked out, I think, in many cases pretty well. Definitely were some points of contention in terms of just making sure that, you know, we were, we were doing a good job of

designing it. Again, when we had to make differences. There's one paper, this was the exosome paper, there was, I think, a lot of back and forth. You're going to read the peer review comments on it. And that's just because, you know, at some point, you do have to be very conscious of technical changes, if we change something, maybe it does make a difference. That paper was one, where they were very nervous about some of the changes that were being introduced. So, we just pulled it back, and got closer and closer to the original methodology. So that's probably the best example of it. Really having a heavy influence on making sure that our methodology was aligned with the original on key aspects. But for the most part, I think it's more about us just laying it out as thoroughly as we can. And that's the beauty of it. Because then, when you follow through with it, and you're like alright, cool, now I can do these experiments. **And sometimes they worked great. They worked like designed. Many times they didn't.** And we're like – that did not work the way that we were supposed to do it. We have to figure out what to do, we'd have to start changing things. Sometimes, we'd go back to original authors, if they engage with us and saying – this isn't working. Can you help us figure this out?

Eric
What do you mean it wasn't working?

Tim
Yeah, sometimes it'd be mundane, those ones weren't upsetting. There's, there's a couple of ones, we give anecdotes to those in the paper. So, I'll give you one, I'll give you two good examples.

So, one is a model where we were making a, taking a glioblastoma cell line, and we're transducing it so that we'd have a stable cell line expressing a certain marker. And we had a GFP expression to dictate if we had enough. And so, got everything from the original authors, the cell lines, the plasmids, remade everything according to their specs. Actually, yeah, they showed us everything, sorry**.** But we couldn't get it to work. We couldn't get the, I think they set it at 80%, they want an 80% efficiency. So, 80% of the cells of the population had to be GFP in order for us to proceed. And so, we were like – it's not working. We're like – what should we do? What do you think we should do? They said – ha, that's funny. Let me send you the plasmid again. So, we got a new plasmid, tried it again. Same thing, not getting it there, just kind of fluctuate. We said – did you ever enrich this? They're like – no. But do you think we can enrich it? Can we use FACS? Can we just sort these out and maybe see if we can grow the population out there? Like that sounds good. We tried that, wouldn't hold. We were like – we don't know what's going on. We're like really confused. Like, we know, we can't proceed unless we get enough cells of this population with that marker expressed. But we're running into challenges. And eventually, we just stopped it. And we stopped it because we just, at some point, we just didn't have a super obvious, I'm sure if we kept trying, we would have maybe figured out a solution. But we were just spending resources trying to figure this out. We just kept experimenting over and over again being like – something's not right. We don't know how to get this to work. But we also know we're not going to go to the animal model injecting these mice until we reach your threshold of 80% of the population. So that's a good example of where we just ran into challenges. Couldn't figure it out. Sounds like it's technical, but really confused because we're using all their reagents and their methodology. So just couldn't figure it out, we stopped.

There's another one, where we had a mouse model. Same thing, and this is a disease one. Everything looked fine *in vitro*, this was the stable cell that the authors gave us. So, it didn't surprise us. So everything looks good. We go through the mouse model, inject it, it's a, I think, it's a leukemia model, if I remember right. And we're going to do *in vivo* imaging, right. IVIS imaging of the mice to check progression, and then we're waiting for the disease to present in all the animals before we randomized to the different drug treatments, either control or this new

drug. And kept scanning, I think, we're like – this isn't working and taking a lot longer than they originally had. And eventually, we started seeing some of the mice progress with the disease before we could even detect it. And we're like, okay, so clearly, we injected them with the cells because the mice are getting the disease. So, we see the onset, but we can't see the abundance in the animal. We're like, there's something technically wrong going on here. So, go back to original authors – what are your suggestions? They basically gave us some suggestions to kind of, you know, tinker around with the lines a little bit, see if we can enrich it a bit more. Did that, did some pilots. Realize that, like alright, we just had to modify it, enrich the cells more, change our timeline completely. It was progressing much longer than in the original paper. Worked with them to basically say – okay, well shoot, like, you know, they want us to treat for three weeks before, we obviously have to sacrifice the animals because of disease onset. So, we tried our best to get that. We're like, this is a tiny window. I don't really know how they pulled it off. But we'll try our best window, based on a pilot now. We did the experiment. And then we're able to kind of get some outcome results from it. Basically, get the model to work so we could detect it. Inject them with the drugs and see what was going on. That took a lot of investment to basically figure it out. hat's interesting in that one is that disease, that model, that animal model onset from when you inject them with these cancer cells, and when you actually see disease onset, when you look in the literature, just looking at the control condition, it's all over the place, right? Nobody talks about that. They're just like – what was my difference in experiment to control? But nobody really always talks about how robust is my model to different environment conditions. And to me, that's actually the bigger story there. It's like, when you actually look at what the original paper did, they were on a super short side, they were saying, inject the mice and days later, you'd be able to see enough of the tumor burden to do something with it. Ours was on the long end, weeks. And there were others in the literature that were as long or even longer than ours. And we're just like – oh my gosh, this is a really interesting model, right? Like the disease model itself that we're using, this animal model in the cell line varies a lot within the literature. And I think that's something to really, not to be upset about, like, that's part of the research, but I think that's something to remember. Which is, those models we're using, the controls that we have, are just as important as the experimental conditions that we're testing.

Eric
It's also interesting that when you're being kind of the explorer, the scientist that's going out and trying to find anything, there's so many challenges, and there's such an emphasis on what comes back there. But it's pretty clear that the reproducibility study really focuses on kind of a different, like it, if someone goes out and does something extreme, you come back and you're like – how do we stabilize this concept? How do we make it really reproducible, really robust? And so, it's almost like the reproducibility study doesn't just give more accuracy, but it kind of fills it out and makes the system notably more robust. Not just – could we replicate it? But in trying to replicate this thing, you'll be really annoyed about the fact that the protocol wasn't there. Like the original author was annoyed that nobody in the world knew about this beforehand, or not annoyed, but like, they, no one knew. But they didn't have that annoyance that they were trying to follow a protocol that wasn't written down well. So like, you're kind of the first person to ever really have that pain. And so, you're also going to be much more motivated to make sure no one has to ever undergo that again.

Tim
Yeah, no. And you know, and you're completely right, which is the way I think about it is, our job, right, when we're replicating it, one of the things that comes out is trying to understand what is responsible for that finding, right? Like, right, when you're the first one that's publishing it, you're like – okay, I think I've got it. And you are trying to describe what you think is responsible for what the finding means, which is important. But we're also saying – how did you get that? Like, what are the conditions necessary to produce that, right? Like, it's not, it's clearly not just mouse model, cell line, use your favorite protocol. It doesn't work that way,

right? There's a lot more in there. And that means there's a lot more variables that are at play, that are really important, so we understand, well – what is this? What is this actual finding that you're having? Can we get it again? And do we actually understand the conditions? And to me, I think the best way to do that is this independence, right? You need somebody else to go about doing it. And the reason why is – I'll do an experiment once in the lab and then I'm like, I'll do it again. I have tons of implicit knowledge in my head, right? It's like, it's like, almost like, I just recall it and like becoming a robot. And like, I know, I'll do everything the same way. And then you have the lore, stand on your left foot, I swear, it only works when you stand on your left foot. But when you have somebody else do it, they're like – why are you doing things that way? Like, like, I don't understand, like, what, what is important? Why are you standing on your left foot, right? And I think that independence is really important. And then I think you're right, what comes along for that is, hopefully, increased transparency of what is needed. And that's the beauty of when we, sometimes it's easy to focus on a project like this, when our replication doesn't give you the same result and say – why did you not get the same result? What's different? There's tons of reasons we can explain it, we can hypothesize what's wrong, sometimes technical, sometimes something biological is going on. We have to remember, the reverse is true when you get the same finding. If you did an experiment and I do it by myself again, and we get the same thing. We did not do it the same way. There's tons of things that you and I changed. And it was robust to that. And I think that's the beauty of it, which is, isn't that exactly what we want. You and I should be able to change things that we don't think that matter. And it doesn't matter, I get a similar finding. So that's the beauty of replication.

Eric
It seems like it goes, I never think about the *Nature* paper as being preliminary data. But the kind of concept here is that it's saying – oh, this might be true. And then in the replication of it, it's in my mind when it would then become robust. Or I definitely liked the idea historically, that a long, major paper from an excellent lab is something I could anchor to. But I maybe have to reframe it a little as its preliminary, for no maliciousness possibly. But just if I wanted to go and do it, either, I would need to look at the methods or maybe there was some, yeah until it's been reproduced, it's really still preliminary, well, no matter what journal it's in.

So, one question I have from this whole process, like, you know, in your paper, you speak a lot more about some of the difficulties, like pretty much none, I think zero of all of the experiments needed, or rather 100% of all the experiments needed some amount of clarification, or four out of 193 of the experiments had the data needed to calculate effect size and power of analysis. And, and so there's a lot of other things we could discuss there. But one thing I wanted to ask about was the kind of, it seems a little bit dangerous. And you talk about it, that people can frame this as, you're out there, and you have people's quote unquote 'careers'. I mean, not their actual careers are affected by this. And so beyond being for science, your work ends up being, having, I guess all science and all publications have some amount of politics, like they get us further in the career, they get us the next level, they get us more funding, but yours definitely has this deeply built into it. Politics, that critique of the system, and understanding of the system. And, but nevertheless, science is often presented as being data centered. Can you, can you talk about some of the experience of interacting with the people how this worked, the emotional responses, the politics of how it worked?

Tim
Yeah, yeah, I get your question. It, yeah, no, it's a good one. I get your question. And, and I'll be careful how I present it because I think I've learned over time how to be more, more careful. So, there's some approaches that I'll start with, my, just kind of some approaches that we took, and how that builds in, and give you some examples of what we experienced. The approach I took is always kind of twofold. One, don't assume, don't assume anything. Don't assume ill intent, even if you get it, like we tried our hardest to be really careful about how we were framing things. It's so easy for anyone, especially because I think something like replication is

perceived the way that you had it, still is now as – oh, you what, you don't trust me? Instead of – oh, that's great, you're replicating it. I think still a lot of confusion in the community about this being kind of more of a threat. Somebody is questioning your integrity as a researcher and questioning your finding, instead of actually being excited about it. Because of that we'd be very careful what we word it because every single, you can imagine, anything I write, even if my intention is good, if I'm writing an email, somebody might interpret it completely backwards. Related to that we always said – we're gonna write every email as if it could go public. Not to suggest, we would publish our emails. But we want to make it that clear, which is, it's not us versus them. It's us being part of trying to understand this and trying to do it transparently. And so if we can't sit there and stand behind our emails, what good is this, right? Otherwise, it becomes a bit of a tit for tat. And we just wanted to avoid that. So we got very good at that over time. Learning never ever to send an email without proofreading, I think it's just a good rule of thumb, unless you're just emailing somebody casually.

So, then the question is, what kind of responses would I get? So, there's, it's really interesting, we'd get it kind of all over the place. So, we can start with the original authors, because there's also some interesting responses from reviewers and editors, as well. So, original authors, I'd see everything. I'd see, got definitely plenty of emails that are saying, this is cool. This is great. You're doing it, I think it's awesome. Thank you for like, I get why you selected these papers. I'm actually very happy to have my paper in there. That's really cool, right? That basically is an acknowledgement your papers is of high impact, right? It was making a noise. I think some people recognize that as a compliment. They thought this project was great. And they were very open and helpful to us, some incredibly helpful. And actually illustrates the challenge, right, which is, my goodness, you had it, that's how the only way we could do it is if you were that helpful. And so that's something to me, I'm like – oh, that's something to fix. Like that burden doesn't fall on you. But it just happens to fall on you. But then I'd see the reverse. I'd see the complete other side of it, where I'd get emails saying this is horrible what you're doing, this is going, this is the worst thing for science. Some people saying – you know what, I'd help you if you weren't doing a replication. But if you're doing your replication, I'm not going to help you. So, we'd get that real, like opposite response of this is the worst thing you can do. And how dare you question what I'm doing?

Eric
Did it work for them when they were being unhelpful? Did it mostly protect them from being investigated? Like were you not able to do those replication studies? And subsequently, those authors were safer?

Tim
That's a really good question. So, I don't have data on this. It's all anecdotal what I'm about to say. It's mixed. There are definitely a couple that were very aggressive in that sense. And when you are aggressive, and you choose not to share something, you can stop us. Absolutely. Because you put the burden back on us, which is, if I have to remake something, because you're not going to share it, the reagents are what stop you, right? Like the data can't stop you, if you're trying to do reproduction. I can probably fill in the methods, I'll probably mess it up. But I can fill in the methods because I can make assumptions like anyone else. But to recreate the system, to recreate the mouse model, that's a cost I have to take on now. And so, there were some people that by doing that, they absolutely stopped us. There was no way that we were going to spend our resources just to remake a mouse model, because you chose not to share it with us. Instead, we said it's better for us to do more and to do it well than to spend all of our resources or the vast majority trying to recreate something somebody chose not to share with us.

Eric
I mean, it also speaks to how masked what was actually happening is if they're completely unhelpful, it's the end.

Tim
Yeah, yeah. Yeah. Which is something to think about. I think that's something really important to think about, because I, you know, as scientists, we, I wouldn't want that to be our perception. And the fact that there, you know, some people, I think, do view it that way, which is science is meant to be very open, sharing what we can, within reason. There's nothing that we were asking for that couldn't be shared within reason, as far as we were aware. And we weren't being told – I can't share you because it's proprietary. We're being told, – no, because you're doing a replication, that's why.

Eric
Usually, the Journal says, – you have to be able to provide this.

Tim
Yeah.

Eric
Is there any recourse? Or just in terms of energy and politics, you didn't want to go down that road? Or is that just…

Tim
That one. It was that one. We chose, so we had a couple in the beginning, we were treading lightly and chose not to pursue anything. We could have. We could have gone to the funder and to the journal. Both of which had policies that you were supposed to share, but we chose not to. And that was just partly because we just knew, what are we going, what's to win there, right? What is to win there? Like is that, is that really what we're setting up this project to do? Instead to us we're like, well, that the project actually, or we've already gotten our answer. Our answer is you don't want to share, unless we come down heavy handed. And is that the best way to go for it? So instead, what we'd have is somebody might say – no, I won't share with you. And then it might be, say a plasmid, which still has cost but it's something that we can absorb. So, it increases the time and cost, ah jeez, we're just gonna make it again. But we know we can do it. And when we do it, then we'll deposit it. Like the next person they can, they don't have to ask us. Go to Addgene, go somewhere else, we'll deposit it for the next person. But just because somebody chooses not to share with us if we could create it, we would just do it again, and we continue on our merry way. Knowing again that impeded us, right? It meant that we had to spend time and resources remaking something when we didn't have to. We could have easily have gotten it from the author or better yet, without even having to ask them from companies like Addgene or you know, ATCC cell lines. Like there are, almost all these repositories exist that can be utilized. So that was the path we took. So, the answer is yes and no, right? Yes, sometimes they stopped us. No, sometimes we wouldn't let them stop us. And part of it was this decision that we had to make. But the annoying thing is it happened. I think that's to me the bigger story is that it happened. And it does happen.

Eric
And it was substantial. It was like 30, something percent. Had a strong…

Tim
Yeah, about a third did either not respond to us. Some just wouldn't respond to us. We knew we had the emails, just wouldn't respond to us. And I think it's, it is what it is. Or they'd say – no, I'm not helping you. There was one, one email to be quite clear in that where they could,

this is another part of it, which is how much of them were saying – no, I won't help you because of what you're doing. There's one, there's one where somebody said – no, I won't help you, because I think it's best you do it yourself. And that one I respect. That's not helpful, but at least you told me your reasons, which is, you think I should do it completely by the paper. So that was actually the one exception of like, somebody actually, I think, thinking hard about the project, 'cause we've never actually technically had to reach out to the authors. You can imagine doing this project without talking to them. And in some cases, to me, science should operate that way. I mean, why are we contingent upon individuals?

Eric
I mean, it is also an interesting concept, what the, as a separate question, let's do a replication study where we don't ask them. And then say – what does these journals that are supposed to at least have the idea of being able to publish reproducible research. If you try and reproduce it just from what the journal publishes, and 0% of them can be, or whatever, then that's somehow also a very interesting landmark concept, that no journal is publishing reproducible methodology sections.

So just a pivot one time. There, as I said, there was a number of issues throughout that you kind of ran into little blocks, and you're trying to work through it. But having spent, with a big group and resources, a number of years, I assume you've also thought like – man, if we just did it this way, it would be so much easier. So, I would like to kind of pivot towards and especially start with, what are some of the things that you think would be relatively easy changes that would enact a big effect for people, like you or you as a PhD student that is trying to reproduce this thing in a lab to build off?

Tim
Yeah, those are two, those are great questions. So, the first one, but like, what low hanging fruit we can do? And it gets, I think, to this big question to have like these are from the past, does it not occur now is not important? I think, I think it still matters. And the reason it still matters is, these practices aren't system wide. They're not, they're not the norm. You look at papers today, they kind of look the same, methods are not great, sharing is not great. Okay. So, doesn't by the way, that's a big statement, doesn't mean that there's not individuals and papers that are not knockout amazing in terms of transparency, openness, and sharing. It's just that I don't think it's widespread yet. So, then the question is, like – how do we do that? You know, one thing that I think is, there's two big things actually. One is, just deposit things, like there are so many, like, resources. When you make a resource, unless there's some reason to really like hold on to it, like deposit it. It doesn't take that much time, and then that burdens off with an organization that loves it, right? If you want to make a plasmid, give it to Addgene, they love it. They're amazing at what they do, right? Cell lines, animal models, antibodies, there's tons of groups out there that really pride themselves in doing a stand up job on that. So, I think you know, a little bit of investment there means that you enable research, and you can track your progress in a way that you wouldn't have done otherwise. And then you haven't have to worry about people asking you. Or you set it up. You can also do it yourself. So that's one, which is try to share what you can.

But I think, the bigger thing that I've really thought through a lot on this, is the methods. How do we make the methods better? And there are some great tools that are out there but there's an approach or a mindset that I took, we took with the project, and then if I had to go back and do it, as my big recommendation to someone else, which is, plan to share from the beginning. And then don't use your methods in the paper to describe everything. It's the worst way, a bad idea, right? So, it's not to say that our methods sections in papers are not good, I think they're good. They're a good way to summarize what you did. But they're not a good way to describe exactly what you did. If you want to describe exactly what you did, show me what you actually did. And so, to me, we live in a digital age, even if you use paper notebooks, you can take a

picture, and you can upload it as a PDF. So, I'm a firm fan, which is if I, when we do this where it makes sense, you know, summarize your methods in the paper, and then put a link to the actual protocol. So, if somebody wants to find it, they can actually see it and if they want, you know, if you're comfortable with that, yeah, take a picture of your handwritten notebook, they can see your handwriting. They'll figure out the details just like you will.

Eric
Do you mean like a protocols.io, a longer protocol or do you mean like? What do you, so that when a person's reading the paper they can summarize it, get a general sense of how the experiment was carried out, but if they want to reproduce it there is the actual kind of code of what actually goes into it. And do you have any thoughts of if, I mean, if a PhD student, or really anybody is thinking about this now, about for their next project, what does this actually look like?

Tim
Yeah. So I think if you know, if you want to use a tool like protocols.io, go for it. It's the best way to do it. If you're doing ground up, I think those tools are amazing because you can share the protocol, get feedback with the community, you can get the version. I think that's a great way to do it, because it's digitalized. I think that if you're not comfortable with that yet, or you're in the middle of something, I literally think that the best way to do it is when it comes time to write your paper and write your methods, just make sure that you deposit like in a data repository, like we use OSF, Figshare does it. There's tons of repositories out there, upload a file of your protocol, the whole thing. The more you can share the better. Not just what you did, like everything that came along, I would upload it all. And then I'd link to that, I'd say here's my method summary, here's a couple paragraphs describing what I believe are the key parts but more than likely I'm missing something that might be important or might not be important. And so, the best way to do that is follow this link. And that'll take you to my real protocol.

Eric
So, OSF also has a place that you can give it an ID number or that you can put it in a database that is going to live?

Tim
Yeah, yeah, yep. OSF does it too. So, you know, in OSF, we did that for our project. And we would upload protocol files to just get a simple short URL. We'd link to the file and say – here you go, this is the methods we did. Or you could link to a whole project that's full of method files. Just like you would for data, right? Just treating methods like data. That's another way to think about it, right? Which is, I upload a bunch of data files, here's all my CSVs, I described them. To me, it's the same thing, here's all my method files. I'll upload all the files, as long as I describe them sufficiently, you can go back and reconstruct it. Now, how many people will do that? I don't know. I don't know how many people will do that. But I can tell you, we did it, we were looking for it. And I think others will. And I know for my PhD every once in a while I'd keep going down the rabbit hole. I'd say this looks really close. I need to figure out their methodology. And then I'd get to the paper's methodology, and it would be unsatisfying. And I, what I really wanted to do was see their notebook. I wanted to say – what did you actually do?. And, and right now, the only way to do that is to email the author. But we know not every author is going to respond or wants to respond, or has time to respond. So, I think we need to get away from that system.

Eric

So, what it is, I mean, I really liked this idea that that is an incredible amount of data in space, that actually doesn't make sense in the method section of a paper when you actually just want to get a general sense. But to say actually, the method section of the paper is a useful function. But now we need to fork that and say there's two functions. One is for reading the paper, and one is for going to the bench. And that when we go to the bench, we need, like a longer repository, and it kind of merges with the idea of, like, consistency in the lab, sharing it with yourself, like a good note taking abilities. It's not like it's a stand-alone concept that's an added burden. But I think I mean, as people use electronic notebooks and stuff that it will be a natural thing that labs, like GitHub or something, that labs have their own internal protocols. And then it's actually not as big of a burden for a journal or someone to say, please put the proper protocol. Yeah, it's a little tricky, because, you know, sometimes people get a second paper out of writing the major protocols for it, but I do, I do think that that's actually a really, a nice, yeah, a really nice idea.

Tim
Yeah, yeah. I mean, it's akin to exactly what you're getting at. Like anyone, not, not everyone does, right? But for those people that are comfortable, and do coding, and are comfortable sharing their scripts, it's like what GitHub did, right? It's saying – yeah, in the paper, I'll report what came out of my code. Of course, right? That's what the paper is for. And maybe I'll describe a little bit of my approach, my methodology to it. If you really want to know what I did, go to my code. My code will tell you what I did, because that's actually the scholarship. That's the only way to go down. And those individuals that really want to that's what they're learning the most. And of course, you can then repurpose them. And again, there's nothing against like, you know, if a *Nature methods* paper makes sense, like, yeah, then great. Like, it's almost like, we don't do that for everything. So, when it makes sense, of course, go that route. I think that's a good idea. But since most of the methodology we do is quote unquote ‘standard technique’, you have a lab notebook, just open it up when you're ready, but only when you're ready. And I think that also goes true for what you tried. I'm a big fan of that. I think of, again this comes a lot from my grad work, we did it a bit in the project. But sometimes we try things, I think, antibodies is the best example. You know, you're like – oh, I need an antibody to detect this, and which antibody should I go? And you go to the literature, and you're very unsatisfied. And then you start trying. And you try your controls, and then maybe you try five. And then the one works, right? You're like – this one is technically good. This antibody does detect my signal the way I want it to work. And that's the only thing we report. Even if you do a good job, we report the antibody and the methods that work. But what we don't report is, you know – I tried four other ones, and they didn't give me anything. And then we keep that silent. And then what happens is the next person does the same thing. And then what we do is we allow this, these antibodies, that some of them are just garbage, or some of them only work in a very narrow sense, and nobody knows about it. And we're wasting, we're just repeating ourselves over and over again.

Eric
Like, you're going to get a lot of criticism, if you say, this doesn't work. Then if you're trying to publish it, then they'd be like – well, you need all the controls. But what you're kind of suggesting is almost like an appendix commentary that is not reviewed. It’s just like the author's post mortem, where they say – I tried these and these. And nobody sees it as factual, but it's just, I mean, it's like reading someone's dissertation or something where you get a little bit of extra. And so you're saying there could almost be a space that is not considered peer reviewed in the paper, but like a little appendix commentary?

Tim

Sure, I would link out again. That's what we do, right? I would do the same thing. I'd say in my methods, I'd say, we tried, we basically say like – we use this antibody, here's a link to the other four attempts that we tried.

Eric
So, you're saying in the linked-out protocols, this is also because it's a non peer reviewed space. It also just naturally opens up to this sort of commentary.

Tim
I would. I think it's brilliant, because I think what it does, it's saying, what actually, what did you actually try to get to the result that you found? And again, I think you're absolutely right, which is, it doesn't mean that, it's like, it could just be a technical issue. It could be the way I was choosing to use it, it doesn't like, it doesn't have to be. But what you've just done is, you've opened the door to what you did, and you let somebody else see it, which is impossible otherwise. Like, how else are they ever going to know you tried four antibodies, and only one worked for the condition that you wanted? And what you found, again, it only makes sense in some cases, but actually, the antibody one, like just it's amazing, it's such an industry, right? There's a lot of money that goes into it. And I really think that there's a number of antibodies that just proliferated around. You know, hundreds of dollars a pop, that many of us know probably shouldn't work, but not enough of us know it, because none of it, nobody shares. Or that it could work under very narrow conditions. And that's once you figure out those conditions, you're good to go. And that, we don't really share that as well.

Eric
So, to pivot slightly, you kind of open up your intro of one of your papers by talking about how science and research needs to be effective at correcting errors to kind of be the self-correcting system as we see it. Do you think that this replication study of landmark papers is the best way forward? Or what do you think is, if we say – okay, we're clearly not efficiently correcting, there's a lot of errors in what you, at least in the parts you were able to look at. What do you think is a good way forward? And is it just more replication studies? Or is there another piece of the puzzle?

Tim
Yeah, no, that's a good one. No, go ahead.

Eric
No, no, or my question is, do replication studies then shine light on where the holes are? And so, then replication studies are the tool? Or are there others, you've already seen that there should be other tools implemented to create a more effectively correcting system?

Tim
Right. So, two big things there. So one is, the correcting system can and is intended to be a broad statement, right? Some people interpret it as correcting in terms of like, oh, retraction. It's not just, that's one way to correct, absolutely. But that's not purely, that's a quite big correction. I think there's everything all the way through in terms of accuracy of what's been reported. Error is included in that for better clarity, right? Like replications help, meta analyses help, right? Which is, it's correcting to get a better sense of what's really going on. But I think the more mundane correction that actually does not occur, and really should occur, we ran into it a lot, is just simple human mistakes, for better clarity of what was done. Actually, what's interesting is, the meta analysis paper, the investigating paper that we published, there was a

mistake in it. And somebody found it on Twitter, and sent it to Brian. And I realized that it was just an error in my code. So, then I fixed it, and then sent it back. And *eLife* quickly updated it, put a comment in there to the original version and then, done. That's the type of question I think that we could be doing more of, right? We think of, and we should think of, publications are static, in the sense that they're your best guess at that time with the evidence you have. So, we don't want to sit there and have a publication like live forever. But I do think we need to remember that, like, we're humans, and humans make mistakes, and we forget things. And we're trying our best and that it's better for us to correct that system and make it as accurate of a representation as we can. Even down to the small details within reason, right? And the system doesn't allow it. *eLife* is a unique journal in the sense that I think they have a system to allow it. They even do like, you know, the ERA, right? The executable research article. As another example, make it so that people can interact with it more. Like, I think that there's lots of ways to kind of correct the literature from – oops, that was the wrong antibody I used, let me correct that. To – oops, that whole experiment was wrong. Let me retract the paper, right? Like, I think that that correction needs to live all the way throughout. I think the current system sadly is potentially over focused, just like replications being considered a threat on the ladder, which is – oh, it's all about corrections. It's all about, tell me what's right and what's wrong. And we know, it's so much more nuanced than that. So, I think to us, to truly see a correcting system, we need to embrace the culture of correcting, and then let the tools come along for the ride to enable us. And I think it's that culture piece that's a lot harder for us to kind of come along with. Which is, it's okay because we're humans, and it's our job actually to correct and if we're wrong, yeah, do the ultimate correction and fix it. And I love people that like, say that. They usually do the most amazing corrections. And the retractions are the people that say, I know I need to do this, because it's the best way forward. But I think, I find just as much value for somebody saying – oops, I made a mistake. Let me fix it. So, the next person coming along reading it gets the right representation.

Eric
Yeah, I mean, one, one question also, similarly, is clearly certain, doing these replications has shed certain lights on things, whether it's the best way or like a long conversation, and certain rules put in place by editors or funding agencies. There could be a lot of ways to improve the situation, but clearly, undergoing formalized replication has its utility. And can you speak a little bit about what is happening in terms of formalizing this, outside of your guys' single project? And is there a lot of momentum? What agencies are doing what? And where this stands in terms of the culture shifting into a formalization or acceptance, if there is some sort of formalized acceptance in the utility of replication studies?

Tim
Um, so I think replications make a lot of sense. Obviously, I believe that 'cause I did a bunch of them. But not, not at, you wouldn't want to do it all the time, right? Even if we got the cost, even if we're saying for every experiment, it only costs 10% to do replication, right? Just kind of making up the DARPA number. I don't think you'd want to do it for everything. That's not, that would be swinging the pendulum from all innovation and potentially, you know, pushing the boundaries, not knowing if any of it is on the sturdiest ground or with the best knowledge to let's be overly conservative. Nothing shall ever get published. And nothing shall move unless we replicate every single finding. I think the answer is more in the middle. And it's really about like, why? What value does replication have? So, to us, this project was more about what you were getting at, which is, it's more of a reflection of the system, saying, people are saying there's something going on, other disciplines are saying they have issues, let's do this in the open. So, this project was not meant to look at any paper, it's meant to look at the system. And it has bias. And we have to be nuanced, and how we interpret that because we only did a small, super small slice of the huge, vast preclinical cancer biology paper literature. But it has utility there because it, like it helps us understand more about the system. I think individual replications on papers, if we're really unsure, can give us a lot of more clarity, like we were just

talking about. It can actually shine a light, because we're looking at it from a different way of saying – well, what really is important? That's a great finding. But how do you get that finding, right? Like, are we sure that's what we're finding and we understand the conditions? Because if we don't understand it, and we focus on the results, we could be tricking ourselves, right? It could be an artifact or maybe it's not as big as we think it is because we don't understand how you actually obtained that finding. So that has high utility at that very narrow sense of like, I'm interested in that part of research, and I want to know if that is solid.

To me, the most exciting one, especially in preclinical space is investment. An investment could be from a funder, I'm going to invest more as a private company, I'm gonna invest more. Why do you think Bayer/Amgen actually wrote those papers as they do replication on the front end? They're saying – before we invest further, we want to see if the key findings hold up. And if they don't, we're not going to invest. I think that's great. That's a really good use of replications as it gives you the confidence of whether I should or should not proceed. I think the same thing should be applied and others have advocated for it, which is, it should apply before we go to first in human. Going to human trials is to me an investment. And that to me, replication and more rigorous replication, like multi-site replication, to me is the biggest utility at that stage. I'd rather be wrong in the cells and in the animals than I would in the human, is another way to put that. So, but I don't think and I don't think you have to do it for everything. I think you do it where it makes the most sense. Those key findings, the part where you're like – should I invest more? Why, why am I making the decision, right? And sometimes it's efficacy, sometimes it's the risk, right? Okay, we believe this drug doesn't cause any side effects. Let's really make sure of that. Let's do it, try those animal models again. Let's make sure that we're right before we go on. So, I think that, I think there's more utility there than we as a research enterprise are utilizing. What I don't see, and I wish I could give you a stronger answer, is I don't see investment in there. I see it, you know, you see it, you hear about it in the private sector, like I have friends that are working in Pharma, in pharmaceutical companies. And they all say the same thing. Like, usually we use CROs if we need to, and we replicate key findings, in fact, we really try to beat that up before we bring it in-house. Because we want to make sure that our scientists' time are spent moving products forward. And that, and that they know that like that's, that's worth their investment early on. Spend a little bit of money, see if it holds up good enough with whatever criteria they want. My guess is, it's whatever is based on their interest, maybe it's effect size, maybe it's whatever, before they proceed. That makes a lot of sense. The problem is those results don't find their way out, right? They're kind of stuck there. Nobody knows what's going on.

Eric
It seems very expensive for each company to do individual things, versus I don't know, somehow collectively vote on a number of studies that are high-value. Or even as a government agency saying, we're funding high quality science and then not being willing to invest in that is also a little bit to create, I mean, I guess the idea is high impact science versus robust science is a little bit of a shifted focus.

Tim
Yeah, it's a balancing act. You want to want both. This is why this is like such a hard and important conversation, right? Like, you still want to have those risky, high impact projects. I don't like, I would never want to not have those as well. I think those are valuable. But if that's all we invest in, we have to understand what comes along with it, which is – wow, is that really the best portfolio, because we're not making sure that some of those, as we further invest, are as reliable as we would hope. So, to me, it is really just, it's a hard one to answer. But to me, I think **t**hat pendulum is probably swung a little too far on high impact innovation, positive shiny results and rush, rush, rush. Instead of saying, maybe there's actually a different way of thinking about this portfolio we call 'scientific research as an enterprise'. And that maybe we should invest a little bit more on the verification. And then the question is – where do you put

it? Right now, it's pretty, pretty empty. Both in and kind of where it's done, but probably more importantly, also how it gets out. Which is, again, like this really weird question, which is most journals. I mean, that registered reports format we use for *eLife*, that's a great way to get replications, in my opinion, right? Good designs, makes a lot of sense. And, you know, you get the feedback and the publication if you follow through, because otherwise people just don't want to publish it. They're saying – oh, your replication didn't teach us anything, because you just found the same result. It's not true. It gave us more confidence. It taught us something, or, – oh, the replication didn't get the same result, the replication must have done something wrong. No, no, maybe the original is not describing the conditions, it's helping us understand better. Both of those are valuable. But we need, we need to have some space for that. And we need to have the funding to support it. And I think, you know, there's little things that I see. There are some projects that have come up that get support, you know the Brazilian Reproducibility Initiative. That's a project that gets support. And I think in the Netherlands, they've got some support. So, you're seeing little snippets, but not to where I think you're truly seeing the buy-in that it could occur at that kind of large level. I mean, look at NIH's budget in the US. It's enormous. And there's no calls for replication studies I've ever seen in their funding. NSF, on the other hand, on the social behavioral, has started to do it.

Eric
Yeah. Well, I think that's a nice place to end for now. I guess I could ask one last question. Is there anything that you think is, I don't know, you're like – ah, I always wish people had this, this idea that bounces around in my mind. Is there anything else you want to say?

Tim
I'd say it's a one last thing and it's just really kind of going a little farther on this last discussion we had about this optimization. You know, I don't, I never started this project. And I still believe that now despite a lot of challenges and a lot of not the nicest side of the culture, sadly, that occurred or I viewed during this project, some of those emails I was talking about, I truly believe, especially in the preclinical space, like, we're making amazing advances. Science is amazing. It really is, right? Like every day, I'm blown away by what we can do. And we know it's making an impact on our lives. So, to me, the most important thing is, not stop investing in science. It's completely right. It's just this question of why are we, is this the only way to invest in science, right? There's another way to invest in it, I think. And I find that a really interesting question, right? So, it's not this – oh, we're wasting science. To me, it's more about it's optimization, right? It's like, how do we, another way to say this, we've made amazing progress, despite all the challenges we just described in our paper. And despite the fact that we could only see effect sizes, you know, at 15% the size of the original. What if we could increase that? Imagine what we could do on top of what we're already doing. So, I really love that framing. That's the whole reason I joined the Center. And the whole reason I started this project, I still completely believe that, which is I think, we can keep pushing it and get better and better and better. This, I really want to overemphasize that because sometimes people are like – oh, you're just telling us science is all wrong. And I'm like – no, it's not. It's amazing, just inefficient!

Eric
Feels like it's pointing to the idea that, like we have, we're building new data on preliminary data, and not necessarily like, we don't have a super solid foundation of key experiments for us to actually be proceeding. And that we're kind of building off of, yeah, fragile space. So, thanks, I hope that this, I mean, I feel like there's a chance that there'll be a ripple effect from this work, at least in, hopefully, in the agencies. Maybe we do a follow up in four years. That's just a celebration of like, all these new agencies that you know about that are now starting to do replication studies. But yeah, thank you.

Tim
Thank you.

Doing Science Differently was produced by the SPOKES Think Tank, a group of early career researchers funded by a subproject of the Wellcome Trust Translational Partnership at the QUEST Center at Berlin Institute of Health.

This episode was hosted by Eric Danner and produced by Joachim Fuchs. Credits by me, Christine Knoll. We would also like to thank Constance Holman, Friederike Kohrs, and Tracey Weissgerber for their help running the SPOKES program.

To learn more about this podcast, visit us on Soundcloud at Doing Science Differently Podcast or follow us on Twitter @DoSciDiff. Stay up to date with show notes and future episodes on Spotify, Apple Podcasts, or wherever you get your podcasts online.

We also recommend you check out the QUEST Center for Responsible Research for lots more great content related to Doing Science Differently.

Sound editing and original music from Giorgio Cattaneo. You can hear more of his musical work on his Linktree page, just follow the link (https://linktr.ee/exoplanetrec) in the show notes of this episode. Or visit Giorgio's personal Instagram account at giorgio.sauro. Logo design and social media management from Anita Waltho. You can learn more about her other projects through her Twitter account @Walthoanita.