Full TRANSCRIPT Season 2 Ep 1

Steve Lewis 00:09
Welcome to Speaking of Mol, a podcast series about molecular biology and its trending applications in the life sciences. I'm Steve Lewis, and I'm so excited to kick off season two of our show with a chance to chat with two brilliant fellow Coloradans. Our guest for this episode is Dr. Joey Azofeifa, CEO of Arpeggio Bio. And our guest host today is Angela Koker, regional sales manager from molecular biology and sample prep at Thermo Fisher Scientific. Our conversation today stretches from the applications of artificial intelligence in molecular biology to life as a startup CEO. You're definitely in for a treat. We begin by asking Joey about his academic background and how it led him to his current work in mol bio.

Joey Azofeifa, PhD 01:05
So, I did my PhD at the University of Colorado in Boulder, and I was in a lab that was really focused at the intersection of computer science and molecular biology. I worked with a Professor, Robin Dowell, at the MCDB department. We were really excited about some of these up-and-coming next-generational transcriptomic assays. RNA sequencing, I'm sure many folks are familiar with, but we were interested in kind of newer methods to get at very sensitive readouts on nascent transcription, or pre-mRNA transcription, which I can talk a ton about. And we were also interested in techniques that were scalable. So, we could use transcriptomics across thousands of samples. And during that time, because Robin's lab was so interdisciplinary, they were generating just a ton of data, and I was developing a lot of really interesting machine learning techniques to understand patterns between, you know, gene expression states, and whether the cancer, whether the cell was cancerous or healthy, even predicting what would happen to a cell, you know, based on a small molecule structure. So, lot of kind of like really exciting, you know, computational techniques. And some of that work, we ended up writing a patent on. A few years later, we started Arpeggio. I met my co-founder, Tim Read, who was on the wet lab side in Robin's lab. And then we met our third co-founder, Laura Norris, who is tangentially related to that lab. And we've been off to the races.

Steve Lewis 02:29
Thank you. That's very interesting. And, and then Angela, as co-host today, you lead a team of technical molecular biology product specialists for the Western USA.

Angela Koker 02:41
Yeah, I'm also excited because I did my undergrad with MCDB. So, I'm also a Buff grad, so happy to see that. There's also so much RNA work happening in Boulder County and coming from Dharmacon, I worked with Dharmacon years ago, I just love hearing about up-and-coming companies as well, especially with RNA. I do have a question, though, because you were talking about pre-AI. But obviously, that's evolving quickly. So, what do you see coming next, when it comes to generative AI and how that's helping you do what you do?

Joey Azofeifa, PhD 03:16
Gosh, please stop me if I talk too much on the subject. So, when I was doing my PhD, I was interested in generative methods. Machine learning when I did it, and you know, between when I was doing and a lot between 2012 and 2015, was very discriminatory in classification, right? It was all about like, given an image of a cat, can you know, the computer say is it a cat or not?

Right? There wasn't this idea of can an algorithm generate something new, right? Can it draw from its trained algorithm and create? And I was really interested in that, could we create machine learning techniques that would generate biological data. And if it could generate biological data, then we could have maybe more confidence in the assertions, the techniques we're making. And I also thought that if you're building generative systems, you have a better understanding of the underlying thing you're trying to model. And now yeah, gosh, so we use generative AI in many different components. One of the ways we use it is to generate chemistry. Typically, when you think about drug discovery 15, 20 years ago, it was very brute force. You know, you hired 30 medicinal chemists, they would make a ton of different molecules. You would empirically test if those molecules worked or didn't work, whether that was killing cancer cells, or, you know, leading to less inflammation. And if you got a hit, great; if you did it, you didn't really know why. And now because we have so much training data on what chemicals look like and what chemicals do, we can build machine learning techniques to predict what a chemical will do from its structure but go the opposite direction. If you want an anti-inflammatory or you want any anti-cancer therapy ask the algorithm to dream up a molecule that would, you know adhere to the properties. The computers tell us to make molecules. Sometimes they don't work. Sometimes molecules aren't even real or can't be synthesized. But sometimes it dreams up very interesting molecules and when we make them, they have very interesting properties.

Steve Lewis 05:18
The way I envision it is, it's along the lines of like, you have a target, right, a biomarker that you might be going after, and you know, the spatial geometry of what that looks like. And you build an algorithm that you think will fit, let's say, as a rectangle fits into, let's say, the receptor, right? And then maybe it's not a rectangle, but it's actually a square. And if you cut a corner out of a square, it might fit a little bit better?

Joey Azofeifa, PhD 05:47
The algorithm that I like to use to explain generative AI from a training perspective, is this technique called generative adversarial networks. And basically, the idea is you have two neural nets. You have one neural net that is trying to discriminate—is this picture, you know, a cat or not. And then you have another neural network that's trying to do the opposite. It's trying to create fake data to fake out the discriminator. And so, these two adversarial networks kind of battle against each other. And the discriminator gets really good. And the generator, it gets really good. We use a technique like that for chemistry. And so going to your, to your point about, you know, how do we design small molecules to fit receptors, we definitely do that. The discrimination for us, though, is we use RNA, which is just kind of a weird idea, right? How could you use RNA to do this? And part of, you know, what we studied in graduate school was, again, very sensitive techniques for RNA sequencing and very scalable techniques. So, you know, we don't sell our platform as a service, but we can do RNA seq for close to 300- or 400-fold cheaper than what is available in typical RNA sequencing, library prep kit techniques. But the cool thing is, is that we have looked across, you know, tens of thousands of small molecules now and hundreds of thousands of transcriptomes. And the patterns that we're learning are similar, you know—what does a chemical look like benzene rings methyl group—but the association is not to a receptor, it's to 20,000 mRNAs. And so you know, if you give me a small molecule, I can tell you a lot about which pathways will get upregulated, which RNAs will get repressed, which transition factors will get repressed or activated, and then go the opposite direction. If there is a cellular network or a cellular state that I want to induce, generate a small molecule that would create that event, right? That cellular state.

Steve Lewis 07:44

It's not just one protein hitting one ligand and then making a downstream impact it's a protein can hit a receptor, and then cause a bunch of downstream impacts. Angela, you have a great molecular biology knowledge and background for RNA, DNA as it relates to some of the molecular biology techniques that we see every day within our company. What would that kind of pace of discovery do not just for scientists in the biotech space, but even in drug discovery and academia?

Angela Koker 08:27
Well, it really depends. Joey, because it's very specific kind of is its nascent RNA, but it's like enhancer RNA. So really short, kind of instable, right RNA versus your long reads?

Joey Azofeifa, PhD 08:39
We do both. The two techniques that we use often are basically a three prime end 3' RNA bar coding technique to really drive down the cost of RNA seq, which I can talk a lot about. But then the other technique that we use a lot is nascent RNA sequencing. And that is not just, "Yes, you detect enhancer RNAs," but you also detect, you know, general transcription, pre-mRNA transcription of genes.

Angela Koker 09:04
I guess the question could go to you, then Joey is what do you think, thinking about the price point going down? If there is something that academic startups could benefit from when thinking about the price of sequencing?

Joey Azofeifa, PhD 09:16
Yeah, I mean, so the price of sequencing is I think, is, well, how's this the right way? Like, why do you sequence and it's or why do you sequence deep? Depth is really the, that the huge cost driver of sequencing and if you can sequence that lower depth, you don't need to spend as much money. For us, we have employed a technique called three prime end (3') mRNA barcoding. Basically, the idea is you have a 384-well, a 1,536-well plate of cells, so live cells. And let's say you want in each one of those wells to do an RNA seq experiment. Typically, what you'd have to do is, you know, suck up cells from each of those wells. You know, extract RNA, make libraries individually, and now you have libraries you send out to sequencing. Number one, that library prep assay procedure is terribly long, if you were to do that, for 1,500 libraries, that's, that's crazy. But then secondly, if you're reading out on the entire transcriptome, again, you're going to need to sequence at a very large depth, maybe like 10 million reads. Our kind of workaround to this approach is to before we do the RNA extraction, like you would typically, we basically spike into each well, a barcode that has a poly(T) sequence on the, on the oligo. That poly(T) sequence ligates on the poly(A) regions of mRNAs. And then we can use in-cell RT to basically amplify that oligo with that associated mRNA, and the associated well-level barcode, which means, we can basically pool now all of those 1,536 wells into a single tube and do one library prep. And because we're only reading out on three prime end (3'), poly(A) sites, we don't need to sequence to 10 million reads. We seq, typically we sequence to about a million. So that drastically drives down the cost as well as the fact that we're not doing 1,500 libraries and all the reagent costs associated with that for. So, for both of those reasons, we can really do RNA seq much, much cheaper. And then because it's a very coarse-grained assay, you know, we'll do like 40,000 small molecule screens, get back a hit, and then if that looks good, then we'll follow up with more traditional nascent RNA sequencing, regular RNA sequencing, and much higher depths to kind of validate, you know, were the hits real. But because it's so cheap, now, we can actually, you know, do this across huge libraries of small molecules.

Angela Koker 11:41

Just to clarify nascent, like happens at the time versus steady state, right? So, it’s more of an immediate impact of a drug or something in the environment.

Joey Azofeifa, PhD 11:52
That's right. That's right. Yeah, we love nascent, again, as a secondary. And we're, by the way, we're trying to work on nascent. For this, like, kind of massive barcoding technique as well. But it's more challenging, which I can talk about. But yeah, nascent is immediate transcription. It's like pre-mRNA. The earlier techniques for nascent RNA sequencing were techniques like pro-seq. And that was a lot of what Robin Dowell and a lot of some of my PhD focused on. Basically, what you would do there is you'd have, again, live cells, you would incubate a nucleotide that has a biotin group on it. And as RNA pol II is turning DNA into RNA, your cells will now have some portion of this biotinylated nucleotide, it'll incorporate that into the growing pre-mRNA strand. And then now the end of that pre-mRNA strand has a biotin group on it, you can then do, again, RNA extraction, and then wash it over streptavidin beads. And now you can fish out, basically the population of RNAs that are being immediately transcribed. The value is really like, you’re kind of isolating the faucet from the bathtub. And you can see very, very fast changes in the cell’s transcriptional response to a small molecule or to any kind of environmental stimuli.

Steve Lewis 13:09
Do you do anything with I guess computer vision related to phenotypic changes? You mentioned kind of like the genetic level and then the transcription factor level. And I'm curious about when it does move upstream to the cellular level, do you do any observations in there as well?

Joey Azofeifa, PhD 13:32
Definitely. I mean, I think it really comes… Yeah, as much as like, especially in this like platform biotech era that we're living in, or I'm not sure. We would love to say every, you know, boil everything down to like transcriptomics is the answer. And that's the only thing that we have to do to like make billion-dollar drugs. But it's definitely not the case. The value in transcriptomics is that it's extremely multivariate. And because of the techniques I described, it's so cheap, it really is so cheap. I mean, we're basically doing transcriptomics for sub dollar RNA seq, right? you can't do that with proteomics, or other profiling techniques.

Steve Lewis 14:09
And is that, this is a question for both of you, is that really going to be kind of that exponentiating factor, the miniaturization and reduction of cost in terms of, I guess, really sending this next level of combined machine learning AI techniques with biology? Are those truly the big limiting factors the cost and then being able to use less reagent I guess, as a result of the miniaturization?

Joey Azofeifa, PhD 14:41
I think the only boat for AI companies or machine learning companies is the data, right? And in biotech, it's even that much more important. So, for us, miniaturizing transcriptomics is going to allow us to create truly massive datasets. I mean, what we already have, from my perspective, the world's largest dataset linking chemistry to transcriptomics, and there is no way we could have done that, not only without the miniaturizations of the library prep techniques, and the assay, and the reagents, and the sequencing, but also, I mean, there's a whole world around laboratory automation that makes it so easy. It's really crazy.

Angela Koker 15:21

And specialized instrumentation for like super small, microscopic volumes. I remember it being a hot topic before COVID-19. I feel like it's obviously still a hot topic, but there's been a shift towards spatial, so more into the multiomics again, and spatial transcriptomics, where you're looking at cell-to-cell variability. I feel like there has been a shift at least on the academic side, where you're able to get more of a kind of a spatial map of the transcriptome, but that's kind of what I'm seeing.

Joey Azofeifa, PhD 15:52
Yeah. And we, you know, batch RNA seq is great, because it's deeply interpretable, right? But to your point, Angela, yeah, it's, you can't get cell-to-cell variability, you can't get within a cell, what localization of mRNAs are being expressed, versus not. I think we have used spatial transcriptomics for one program. But because we're able to scale batch RNA seq to the extent that I'm describing, the data kind of looks like single cell except that each you know, if you remember those like TCR, or UMAP plots, except, you know, in those plots each dot is a cell in that case, because we're doing the single cell RNA seq. But in our case, each plot is a separate experiment treated with a separate compound or a separate concentration at separate time points. We have just extremely fine metadata associated with each of those points. It gives us a different dimension. But for sure, spatial transcriptomics is a such a powerful technique.

Steve Lewis 16:50
We hope you're enjoying this episode of Speaking of Mol Bio. We wanted to take a quick moment to remind you about the Invitrogen School of Molecular Biology. It's a great educational hub for molecular biology with rich and reliable technical content designed for new and experienced molecular biologists alike. Check it out today at **thermofisher.com/ismb**. And now back to our conversation with Joey Azofeifa and Angela Koker.

Steve Lewis 17:26
Now, we started this conversation very, very technical, but there's a really interesting pipeline that you've gone through that I think a lot of our listeners would be interested in hearing about. You went PhD. You went to incubator. You went to VC, funded a company. And I think that that's a very unique experience, especially in biotech. I mean, Angela, I'm curious, you interact with a number of startups and have you seen more today and then from there, I'd love to hear more about Joey's story coming from his PhD program and how he got to today.

Angela Koker 18:05
Oh totally. Across the board there's a marrying of technology, and the AI, and the ability to go through all the data. I think Joey called that out. That is like such a massive thing to cover. So absolutely.

Steve Lewis 18:19
And Joey, yeah, I mean, you're in the thick of it. So, I'm curious to hear how you went like, just tell us your story.

Joey Azofeifa, PhD 18:26
So right after graduate school, I worked for a company called Forma Therapeutics. They were a pharma company. And they were making small molecule inhibitors of epigenetic proteins. And I was talking to Forma, and I was like, "Hey, you guys should really consider some of the techniques that I worked on in graduate school. Nascent RNA sequencing, the some of the earlier versions of the scalable RNA seek that we're talking about." And they were basically like, "That sounds really cool. Sounds challenging to license from CU, University of Colorado. But if you ever start a company, let us know we would be your first customer." I guess it was like a

light bulb. And I was like, "Maybe we should really think about that." So, we all kind of started talking and we started the company, we pitched Forma, they ended up being our first customer. And again, back then, it was really very simple. They would send us cells and small molecules, and we would treat cells and we'd look at what changed in the transcriptome, and we'd sell that data back to them. Through that process and then working with so many pharma companies, Tim and I were like, “Gosh, like, it's not like the scientists are doing bad themselves. But it's like, we could make drugs too.” you know, we were like, “We're helping all these folks make drugs. But this technology is, we're the best folks in the world that know how to use this hammer, like, lets us like, why not us use it to build our own drug program.” During the pandemic, we were really trying to use the technology ourselves to find interesting molecules. And we started on a pathway related to iron homeostasis, because we had strong conviction that iron is kind of this untapped well of pathology, both in terms of cancer, but also in terms of diseases like neuro degeneration, which I can talk a lot about that. But we found this really cool molecule that was disrupting iron homeostasis, and it was killing cancer cells, at least on a dish. When we found that molecule, then that kind of spurred our series A. The last year has been extremely cool, because now we've moved our molecules from dish science to *in vivo* pharmacology, where we're actually dosing orally in a mouse that has a cancer on the flank of its, you know, on its back and seeing that cancer go away when it takes our drug.

Steve Lewis 20:31
You know, entrepreneur, right, scientist, I think that is just an incredible journey to be on. And it has been, I would say, growing as a pathway for people coming out of the lab, and then creating their own journey, if you will, right? It’s traditionally, you know, go to undergrad, masters, and then get a master’s along the way through your PhD. And then consider academia for a very long time and go for grants. And more and more, and as Angela alluded to, venture funding is, is in the space, it's expanded to a lot of different firms that are really fascinated by this. One of my old mentors, old bosses, who will actually be on the podcast later in this season said something that I'll always remember is like, "Technology is only cool in the context of biology." But what I think is really interesting is that we have much more computational power than we ever have. And biology is a problem that not necessarily can be solved, although I think there are a number of computer scientists in the valley who do look at a cell and say, "Oh, you just need enough algorithms." And I'll say, for all the tech people who are who are listening, biology is, it's not even orders of magnitude, it's infinitely harder than technology. And, you know, I'm a bio guy and I might be biased, but there hits a certain point. And I'd love to kind of hear both your takes on this when you do talk to entrepreneurs who might be coming from another discipline, and then they hit a point. I have this kind of like, my own personal opinion is like, you might be like the greatest programmer in the world but if you're tackling biology, you hit a point where you have to know molecular biology to some degree. You have to know the central dogma, and then to a degree I also think you have to know cellular biology in order to understand it in context. I don't know I I'd be curious to hear your take. Is that a too aggressive of a stance or do you see it as it's not it's not such a quick translation that somebody technical could get into biology just kind of you know, reading Wikipedia pages or going to what does it get Github just catching up through there?

Angela Koker 23:06
I think we have a funny example with Elizabeth Holmes like talking about Theranos, you know. Like where you have a good idea and you're a technology person with you know, and you think that your idea will go far. I think that's a prime example of not having the biology to back it up. But Joey, I'm curious what you have to say.

Joey Azofeifa, PhD 23:33

Oh, gosh. I mean, I think that is where to begin. So, one of the things, I guess, like reflect on the piece, Steve, you were saying about just like needing to know, the cellular biology and microbiology as well and to kind of get past that, that programming hill. I think it's even more than that. Like it's crazy, at least at a pharma biotech, the sorts of things, at least as a CEO that I have to do, and from a technical side as well as business development side, is absurd. It's not I mean, yes cellular, yes molecular, yes, programming. But we talk now to clinicians, right? What is the sort of drug profile I need such that clinician would be excited about prescribing it, much less it even working in a phase one or phase two, right. So now you got to think about patients, you have to think about like, is there a benefit to providing a safer chemotherapy? Sometimes there is, sometimes there isn't, right? If the markets really flooded for seven target therapies for estrogen receptor–positive breast cancer, maybe that's not the best place to go if you're a nascent, kind of drug discovery company. So, I think a lot about that.

Steve Lewis 24:26
And that's because of how technical it is to get to that application area, right?

Joey Azofeifa, PhD 24:31
Totally. And the other thing that I really appreciate, so kind of going back to this, like, tech, bio, or just tech-enabled biology, or whatever. I think, to your point, it was a little bit about like, all these Silicon Valley CEOs make a ton of money on apps and pets.com, or Uber for pets, or whatever. And it's like, oh, it's going to be so easy to cure, you know, cancer. We'll just apply machine learning techniques and we'll just, we'll just, isn't it isn't it just that easy, just hit enter. And that hubris did not bear out. But one thing that did bear out are an automation around very early drug discovery, right? So this idea of like, screening large libraries, or going back to the earlier part of the conversation, like generative chemistry. I actually believe that AI can, and we use it, you know, ourselves, like can make a big impact on, you know, finding an interesting hit, and maybe progressing that molecule to very, very early, you know, *in vivo* proof-of-concept very early. But then post that, you know, you still need to spend $400 million on a, you know, all the way from IND enabling to phase two, phase three. And so, so yes, AI is helping, but, you know, thank God for the FDA in some ways, because we still have to run rigorous clinical studies that are placebo controlled. And those just cost money. Because, I mean, you're basically asking a patient to come in, be monitored, take a drug, sometimes daily, like, it's just an expensive, it's an expensive endeavor. And so, I don't know, and there are plenty of AI companies that are working on the clinical side, but at the end of the day, I think the biotech is still going to have to pitch their drug program to a VC. And that VC facility has shelled out like $200 million to get over the finish line. It's a harrowing journey.

Steve Lewis 26:28
It's just very interesting. And we're so lucky to be able to speak with somebody like you, who does have the expertise as a computer scientist and also a bench scientist. And I'm sure you're still very involved on the technical side, but I know as the CEO that brings a lot of the business strategic challenges that I'm sure you, you grapple with day-to-to-day, so it's really amazing. Joey, as we're nearing the end of our time, I'm curious, I do have two closing questions. Very first, what has been your secret to success?

Joey Azofeifa, PhD 27:07
Um, I don't know if I'm successful. I mean but thank you for thinking that. I think, for me, one of my curses, I would say as a personality type is that I really, truly cannot rest until something is done. I remember it being this way in graduate school, which is a terrible, terrible phenotype. Where, you know, if the experiment wasn't completed, or the analysis wasn't done, I literally just could not fall asleep until it was finished. And drug discovery is never finished. It's so annoying,

right? I mean, we're all there. We'll get to monkeys and won't be finished. We get to phase one and won't be finished. Phase two won't be finished. We'll sell the company, and they'll get marketed and it won't be finished because we won't know the long-term side effects of it, right? God forbid that even happens. And then you know, I think both my co-founders Tim and Laura share very similar characteristics. We have a hard time letting things be incomplete and for a company I think that's it that is a recipe, some recipe for success. Because, we just keep going you know you keep going and I don't mean to like it all entrepreneurially on this point, but it's like the ups and downs are just are absurd, right? Early employees leave. VCs back out. Pharma partners back out. Pharma partner puts out a paper that says your, the target of your drug is a bad target. We hope that AI and empirical high-throughput technology can make the risk more manageable and more predictable. But we are literally in drug discovery. Like this is, this is, you know, if you're if you kind of oscillate on those ups and downs, you can drive yourself crazy. So, I think the team, again, while I speak the team, I mean, I think about Tim and Laura, we've all had to really kind of exhale as we've matured as founders to, you know, let those ups and downs kind of move through us. But then also kind of understand that we're just not going to be satisfied until it's until it's done.

Angela Koker 29:02
I do have to ask, though, who's the musician with Arpeggio? It's kind of a beautiful name.

Joey Azofeifa, PhD 29:06
Thank you. Yeah. And I played music growing up. And Tim also played music. So. But it was neither of our ideas. We were in Boston and trying to think of ideas for the company. And our friend Lavon was like, "You guys, you know, talk a lot about music and time. And transcription obviously has a time component to it. What about Arpeggio?" And we're like, "Oh, my gosh, that's great." So that's how it came apart.

Angela Koker 29:29
There is something, there's something nice to it, like the concept of a chord that's broken, and like, piecing it together. I don't know, there's oh, cool to it. Yeah.

Joey Azofeifa, PhD 29:37
Absolutely. And that's how like, when we first started the company, we did a lot of kinetic studies, where we would look at, you know, how a cell responds over time in terms of the transcriptome and at very densely, you know, taking snapshots of transcriptomic perturbations every 15 minutes for many hours. And, you know, I've been thinking a lot about this concept of like, I don't know, Tim doesn't like this analogy, but I just can't get out of my head where a cell is like a, it's like a gong. And there are just little individual hammers, and these are all different drugs. And if you watch how that gong vibrates at different spots of your hitting it with the different hammers, you can kind of image the gong just through the sound vibrations, right? And that's kind of what you do with drugs, right. When we take a HeLa cell or a HEPG2 cell or whatever, when we expose it to 10,000 perturbations, and we monitor 20,000 mRNAs and 20,000 potential vibrations, yes, we're learning something about the hammers and the compounds, which is very interesting. But in mass, we're also learning and imaging in a very weird way, the state of that individual cell. I do think there is like a relationship between kinetics and time series and gene expression in biology. Yeah.

Steve Lewis 30:56

That was Dr. Joey Azofeifa, CEO of Arpeggio Bio in Boulder, Colorado, and Angela Koker, regional manager for molecular biology and sample prep at Thermo Fisher Scientific. Thanks so much to Angela for helming this episode with me and to Joey for our great conversation. Speaking of Mol Bio is produced by Matt Ferris, Sarah Briganti, and Matthew Stock. We have an amazing slate of guests lined up for our second season and you won't want to miss them. Make sure you're subscribed to our show wherever you get your podcasts and consider sharing this episode with a friend or colleague who also might enjoy them. Until next time, cheers and good science.