AI-GR Podcast #8 Atul Butte 08.21.23

[00:00:00] So we wanna make sure that we are doing this right by patients. At the same time, though, we are now in the kind of mantra now where if we don't do this, something's wrong with us. Because if we don't use this data to prove the practice of medicine, what a tragedy it will be given the billions we've spent on electronic health records systems.

[00:00:18] So that's the narrative I tend to use in my talks. Again, safely, responsibly, respectfully, but I say this all the time, this is the most expensive data in America. We're literally paying doctors in their pajamas to write these notes. It will be a tragedy if we don't use this data to improve the practice of medicine.

[00:00:39] Welcome to a new episode of NEJM AI Grand Rounds. I'm Raj Manrai, and I'm here with my co-host Andy Beam. We're thrilled to have Professor Atul Butte on the podcast today. Atul is a superstar, biomedical informatics researcher and entrepreneur at UCSF, where he's the Priscilla Chan and Mark Zuckerberg distinguished professor.

[00:00:58] He's also chief data [00:01:00] scientist for the entire University of California Health System. Andy, this conversation was really fun. Atul shared a lot about his journey, both as an academic and as an entrepreneur, but the highlight for me is always his infectious energy. He's such a dynamic and engaging speaker, and I left energized and probably with an elevated heart rate afterwards.

[00:01:18] It was pretty cool that he also shared how he improved his ability to speak and also how it led into a pretty major career transition. I completely agree, Raj and Atul is truly one of a kind. I've been a big fan of his work since the days of my postdoc. I think he always has been the bar that many of us in biomedical informatics aspire to and his influence on the field has been immense.

[00:01:37] If you've ever seen Atul give a public talk before, you know what a gifted public speaker he is, and one thing that I really took away from this conversation is how intentional that aspect of his personality is. He really puts a lot of thought and care into how to convey his research to a broad audience, and we could all learn a lot from Atul in this regard.

[00:01:55] The NEJM AI Grand Rounds podcast is sponsored by Microsoft and [00:02:00] Viz.Ai. We thank them for their support, and with that, we bring you our conversation with Dr. Atul Butte.

[00:02:08] All right, so it is my great pleasure to welcome the man, the myth, the legend. Dr. Atul Butte, thank you for joining us today, Atul, on AI Grand Rounds. Thanks for having me. So, Atul, this is a question we like to get started with and that we ask all of our guests. Could you tell us about the training procedure for your own neural net?

[00:02:26] How did you get interested in AI? What data and experiences led you to where you are today? Great question. So, I've certainly been in this field of research, I guess for more than 20 years. As you know, I trained with our good friend and mentor, Zak Kohane, way back. I picked up a lot of skills from him, but also other mentors from the whole Harvard ecosystem, how they ran their journal clubs.

[00:02:49] How they ran their lab meetings. And so that's how you build up your own neural network, right? You can build it up yourself by reading on your own and learning yourself or bringing in talent and having them present it to [00:03:00] you. And I think that's the right mix. My own kind of AI reinforcement comes from devouring Twitter and now Threads.

[00:03:07] I follow thousands of people, so I kind of don't understand those folks that only follow like a couple hundred. I have thousands. That being said, I don't get to see every tweet from everyone. I don't really care about that, but I do like to see what's on top of the Threads. And I probably save too many things to an app called Pocket, which is to read later.

[00:03:26] My Pocket probably has 10,000 articles, which I'll never read, but I pretend like I've read them, so maybe osmosis. So that's kind of my own, I guess, reinforcement learning. But I also bring in a lot of talent, a lot of postdocs, especially in the early days when I was building my lab. I love to bring in postdocs at new datasets that we had no knowledge of.

[00:03:47] So you had to know some kind of new utility or a new method, but also a new dataset, and that way we could leverage the network effect. So that's perhaps how I do it. So that's a good snapshot of your current gradient update [00:04:00] procedure. Could you tell us about your initial conditions? So you've trained as a computer scientist, you're also a card-carrying clinician.

[00:04:06] Take us back to the initial conditions and, help us understand how you got to where you are. Sure. Initial conditions really go back to eighth grade, ninth grade when my parents were lucky and privileged enough to get me an Apple two plus my brother and myself. So started coding very early eighth grade, ninth grade, still code.

[00:04:24] I had to shut down my coding windows for this podcast, right, to keep enough bandwidth going here, but otherwise they would be open. You'd be seeing me writing code. So I've been coding for a long time. So Apple two plus and a Mac in high school got inside Macintosh, which were the actual books, you know, used to have to learn through books, uh, and learn how to write code for the Mac before I graduated high school.

[00:04:45] So then I got into this program at Brown University and there are many in this field who passed through Brown University, uh, in bioinformatics, including Zak. I was in this program, it was eight years, so not really accelerated necessarily, but you can [00:05:00] major in almost anything you wanted to and still go to medical school.

[00:05:02] So I did computer science as an undergrad, really got deep into coding, including AI, machine learning. These were the late eighties, early nineties, but even took a neural network class back then. Okay. So in, it was actually in cognitive science, but got to learn very early about perceptrons and things like that.

[00:05:19] We just didn't have the compute power at the time. So back then my summer jobs were working at Apple and Microsoft. I was on the finder team at Microsoft back in 89, 90. There were only like nine buildings at Microsoft at the time. And then the subsequent summers and winters, I was at Apple on the finder team, adding all sorts of new features into the pre-Mac OS uh, x uh, finder.

[00:05:40] This was something called System 7 way back, and then went to medical school and a year off at that point to go to NIH and learn some basic biology. Because the Genome Project was starting, I had never really stepped foot into a biology lab and really got my hands really wet with biology, uh, sequencing, pouring sequencing [00:06:00] gels.

[00:06:00] And learned about this interesting condition called leprechaunism, where kids can be born with these mutations in insulin receptors. You can even be born without an insulin receptor. Fascinating to me. With the Genome

Project just starting at that point, I decided I wanted to go into pediatrics and pediatric endocrinology, and guess what?

[00:06:18] Zak is another pediatric endocrinologist, and so I literally spent time one month hanging out at Boston Children's trying to learn who the clinicians were, and eventually got in there into the residency and then fellowship and spent in the end 10 years at Boston Children's learning genomics. It was a microarray revolution with gene expression and all that, and Zak and I were in the right place at the right time.

[00:06:42] We went from like zero to 16 grants in two years and just, uh, the whole thing took off. I'll pause there, see if there's questions on, on that part before I talk about the Stanford case. Yeah, I actually had a follow-up question about that later, but I'm, I'm just wondering if there's something about

[00:06:57] pediatric endocrinology or being a [00:07:00] pediatric endocrinologist that makes you predisposed towards informatics and data-driven approaches? Or is it just coincident that you and Zak both happen to be pediatric endocrinologists? It's, it's not just me and Zak, it's also Dan Nigren who was here at Boston Children's for a long time.

[00:07:12] So, but you're only seeing the few that were there, right? I mean, if you wanted to go into computers and pediatric endocrine, that was really the only place. I mean, uh, there are only about 30 to 40 pediatric endocrinologists trained per year in the entire United States. So it's a very select sample that you guys have seen in the Boston ecosystem.

[00:07:31] I wouldn't say that most go into this. Most people in pediatrics go into clinical care. I think, uh, very few even go into specialists, right? Maybe a thousand a year going to neonatology, which is like our intensive care unit, right? For little itty-bitty babies, but most going into clinical care. So again, Boston Children's is almost one of a kind.

[00:07:52] There are obviously other good pediatric hospitals, mine included here at UCSF. The ones that go into academia, especially this [00:08:00] esoteric thing of computers and informatics. Very, very few people, especially in pediatrics, going to, you just happen to see them all there and have passed through those. Got it.

[00:08:08] Those walls. Got it. Thanks. So Atul, I love that you started your personal story with mentoring and uh, some of your mentors that you had along key parts of your journey in your training. So I think that has really come

through in the way your mentees and your trainees speak about you and how important mentoring is for you now, uh, as a PI and as a director of a lab.

[00:08:31] And while I wasn't in your lab, I have to say that I think even you had a large and memorable impact on me. I think this was the first time we met. I was an undergrad at PSB, the Pacific Symposium on Biocomputing. I was in Shirley Liu's lab, and I think Shirley had just graduated from the Stanford program.

[00:08:48] And you really took the time at the conference. You know, I was a very young undergrad. You spent a lot of time telling me about grad school, inviting me to dinner with folks at Stanford, and really, I [00:09:00] think getting me amazingly energized to go into this career and to do academia. And I have to say, I, I really appreciate that personally.

[00:09:07] And, and I think, you know, Turag is in our department, many other folks speak very, very highly about how seriously you take mentoring and, and how much you've had a lasting impact on your students. Well, I appreciate that. I, I, it's, it's also funny you were saying, I take it seriously because I think it's the fact that I take it less seriously that maybe makes it more useful.

[00:09:27] It's a pretty loose lab here. I mean, people, if you ask people, I hope they say it's fun more than anything, right? I don't send emails on weekends. I don't want a response on weekends. I might write them on weekends, but they're timed to go out Monday at 6:00 AM so it's on the top of your inbox. I don't want to deal with your response on weekend.

[00:09:47] You know? So like, I think it's, it's not like we're working crazy hours. It's, it shouldn't be about that or we're doing something wrong. Again, I learned from others. Zak is one of my mentors, I have to say. Uh, Ron Khan, who was [00:10:00] the president of the Joslin Diabetes Center, was another great mentor of mine. I had others during the NIH years, Simeon Taylor, whose lab I was in.

[00:10:08] So you learn from these folks, you know, uh, along the way. Michael Kwan was another one who was a postdoc I was with at NIH. And even the managers I had at Apple and even Microsoft, I still remember them and how they managed their teams. So it's really, you have to respect the folks in your lab, first and foremost.

[00:10:26] They do not need to be working for you. They definitely don't need to be helping you. And academia is the kind of thing where you need their help

to write papers, to do the work, and you need their trust. And you have to trust them, right? Especially in today's complicated world where we have complicated experiments, complicated code, it's not the case where the principal investigators checking every single last line of code from every postdoc.

[00:10:49] So that trust has to be there and trust, meaning like if you write all this code and if it doesn't work, I'm not gonna yell at you. I'm not yelling at anyone in my lab. I don't want them to fake [00:11:00] data. I don't want 'em to kind of do these things that we've been hearing about the last few weeks especially.

[00:11:04] So you have to earn that trust and you have to communicate that trust. So, we do very few things in the lab. We do a lot of social activities, probably fewer now than we used to, but maybe I'll pick up a little bit. Especially post-Covid. I have some expectations for folks in the lab, but we also make it all about publications and publishing if you're in my lab.

[00:11:25] But not in a crazy way. I mean, most people in my lab write a paper a year. So when you have about 15, 16 people in the lab, that comes out to our average. We publish a paper in my lab about every 18 days, and that has been our run rate for about more than 10 years, I think. And so everything, every meeting is about the paper, right?

[00:11:45] What's the next paper you're writing? Is this of interest? I can go into more details how I run my lab, otherwise, uh, you let me know. I maybe have a question about that. And I think it's a generalizable question to a lot of other domains. I think it's been the experience for [00:12:00] most of us as we progress through our career, rather in academia or industry that you end up managing people.

[00:12:04] Yeah. And I think one of the challenges as you grow into those management roles is context switching. That you have to be able to quickly hop into a project, assess the strengths and weaknesses, and then provide feedback to that. So do you have any lessons? Your lab is large, you've been doing this for a long time, clearly you're successful.

[00:12:20] How do you manage the context switching and making sure that those individuals are getting feedback that they need? Yeah, so these are important questions and again, uh, believe me, I don't do this perfectly nor would I ever claim to be doing this at all, even close to perfect. One style thing for my lab is every single person in my lab is on a completely separate project.

[00:12:41] So there are a lot of folks, especially in informatics by informatics, that have one big project that everyone's contributing to. What I saw in the, especially training in the Harvard ecosystem is that makes things too complicated. Who's writing the paper? Is it me or is it you? And then you start to get internal fights.

[00:12:57] Okay, I need a starred second author within the [00:13:00] lab. You know, that kind of thing is complicated. So I started from day zero. Everyone's on their own project, so everyone in the lab, if I point to anyone in a lab, they really should be able to tell me what is the paper they're writing right now. Okay. That they are the first author of – no question.

[00:13:19] Now, what that means is if you want middle authorships, you have to help everyone else in the lab, okay? It can't be at the detriment of your own project, but you have to help others to become a middle author. And so we maximize that as much as we can, but you shall never feel threatened about who's writing your paper because it's your project.

[00:13:37] Just that is 99% of the drama in many people's labs that is eliminated because everyone knows they're comfortable with their own project, their own paper, and then everyone helps each other. And, uh, we get papers out the door. Does that make sense? I mean, that's, that's one sense of drama that we try to eliminate.

[00:13:56] Yeah, I think clearly defined roles as a credit assignment [00:14:00] mechanism, right? Like that's, that's the, the key challenge in academia. I think that makes total sense. Yes. And that's where you get to the credit assignment problem. Hold on. And that's the sign I have here, which I learned from Zak. And you can't see this on the video, uh, because there is no video here.

[00:14:14] But this is, uh, I mean, Zak says, Pete Salvage said this, uh, this quote comes from Don Berwick. It says, "Credit is infinitely divisible. Give it away every chance you get, and there's always plenty left for you." Right? So the whole, whole point is, and this quote is a perfect example of that, since it's been assigned to so many people.

[00:14:34] Anyway, so I think the, the credit, the, the quote is self-demonstrating the quote. Uh, so can you hold that up again? I wanna get a, I wanna get a screenshot of that side. Okay. It's pretty awesome. Yeah. There we go. Got it. All right. There it goes. Okay. That's a screenshot. I think we got it. Okay, now you got it.

[00:14:50] Yeah. Cool. Yes, that's the poster. And it says share. I'm guessing you're gonna, I'll just download it just in case you didn't get it. [00:15:00] Okay. Yeah. So this is, uh, you got it. Okay. My postdocs, one of my postdocs, Javir Iran, got me this sign and he, he's a believer of that too. So that's another one of those kind of feed forward pieces of advice that is very easy to pass on to your mentees, and then they become mentors and pass it on.

[00:15:17] Again, what does it actually mean, right? If anyone helps, really, they should be a middle author. Why are you fighting over middle author slots? Who cares? If there's too many authors on a paper, if they really helped, you gotta recognize that and give them the credit. That is academia. That's all that sign really means.

[00:15:34] I love it. So, Atul, I think before we jump into your research in detail, I think you were taking us through the chronology of your career. Yes. You were just about to enter the Stanford years. So maybe we can just spend a couple minutes pick up there and then we'll jump into your research. Yeah. So, uh, like many mentees and, you know, junior or junior faculty, I wasn't even in, its, I think I was an instructor at the time in the Harvard ecosystem, but I was mostly at Boston Children's, which is [00:16:00] its own little thing, big thing.

[00:16:02] So at some point you don't even know you're supposed to be looking for a job, right. No one tells you you should be looking for a job. That's something I definitely change in my own lab. I tell people you have one year until they say you're not here anymore, so we gotta start this process. So that's something I kind of do at my lab, but back then, uh, certainly no one told me.

[00:16:20] And Stanford was really starting to raise some money around pediatric research, uh, at the time. Just to give you some context, Brigham and Women's Mass General had just formed a partnership called Partners at the time, not Mass General Brigham. Uh, and I guess people were trying to copy paste that. And over in the West Coast, UCSF and Stanford were thinking they could form a partnership.

[00:16:42] Well, the consultants said it could be done. I guess the culture said it could not, and that was immediately falling apart. But in the meantime, while they were connected, the David and Lucile Packard, uh, foundation, uh, had given money to the Lucile Packard Foundation for Children's Health to fund a huge [00:17:00] pediatric research and clinical development between the two.

[00:17:04] But in the divorce, Stanford seemed to get all the money, and so that was half a billion dollars to build up pediatric research at Stanford. So I was

one, it was complicated at the time, they were trying to get Zak. Zak's stuck around in the Harvard ecosystem. That's gonna be fine. Obviously, he certainly can't say he hasn't, uh, done a lot of thriving there.

[00:17:24] But then I ended up getting that offer and moved to Stanford back in 2005, then moved up the food chain from assistant, to associate, to full professor, and division chief over 10 years. And, so, no regrets. And sometimes, again, I tell that to all my trainees, right? There are these mentors that try to keep people in their lab as long as they can, right?

[00:17:46] I am so the opposite of that, okay? Because again, I, I want to believe in my folks. They cannot thrive being a trainee in my lab. They have got to get a real job and go forth. And [00:18:00] so I try to get people outta my lab as fast as they can, especially if they're going into the industry. If they're going in the industry, there's really no reason to do a postdoc.

[00:18:08] Certainly past a year, I think the pay is better. Everything seems to be better for people in industry. Why would you want to stay at postdoc in my lab? So, I, I don't say that to kick them out because I'm trying to be mean. I'm trying to help them. I think they get that. So I left 2005. I was at Stanford for 2005 to 2015, and ever since I really did okay at Stanford, they gave me a lot of startup funds, including endowments, uh, which were killer, which meant I didn't have to write grants, but I wrote a ton of grants.

[00:18:40] I got a whole bunch funded. It's funny you called me a card-carrying clinician. Certainly not now. I never even applied for my license in California because I thought, you know, if these grants failed, maybe I'll go back and add some clinical time. I didn't have to. Luckily, I was privileged in that way and then took off.

[00:18:58] My lab at the peak was [00:19:00] 25, 30 in the lab at the time. Now we scale back to 15 or so here, but I was there for 10 years and then University of California made me an offer I could not refuse. Despite everything I had at Stanford. They made me offer I could not refuse. Let me pause there to see any questions about the Stanford days?

[00:19:16] I was gonna say that there's, there's a mafia reference in there somewhere, and I, I'm, I'm guessing that the offer goes better places than the Godfather, but oh, it's, uh, you know, the Godfather looks out for you too, right? Uh, it's not all about horseheads in beds, but, uh, yes. Uh, you can read into that, but I love

[00:19:34] the offer. I mean, look, uh, so what is it that got me to leave Stanford? I, I still say this in public. I still say this in private. I never had a bad day at Stanford. Okay? I was literally just one week short of 10 years, didn't mean much of a hit the actual 10 years. I maybe would've gotten a pen or something, but didn't matter.

[00:19:51] But I never had a bad day at Stanford. They treated me incredibly well as a junior faculty. Incredibly well, no doubt about it. But [00:20:00] UCSF is right up the road and UCSF is also making inroads. And it's interesting here because we have the two academic medical centers in the Bay Area. We have UC Davis, but that's further out.

[00:20:10] And we have other kind of for-profit medical schools, which don't really matter much. So, it's really Stanford and UCSF and both are gonna keep poaching from each other forever to kingdom come. Okay? And both are gonna train and kind of poach the other side, which is fine, that's how it is. But UCSF is certainly growing like crazy and I could see that.

[00:20:29] So, UCSF spent four years. Trying to recruit me four years, and it seemed like a lateral move for a long time. And it goes back, uh, again, if this history is of interest, back in, I think it was in 2011, there was a whole effort at the National Academy of Medicine starting about precision medicine. It wasn't even called precision medicine at time.

[00:20:50] It was called personalized medicine. And I was invited, and Zak was part of this, uh, because he was part of the committee. But Keith Yamamoto, who is still at UCSF, [00:21:00] uh, invited me to give the keynote at the National Academy of Medicine thing to launch this whole discussion. And there were others there too.

[00:21:07] Uh, certainly a lot of luminaries today and in the past, trying to figure out what personalized medicine was gonna be about. Sue Desmond-Hellman and Keith Yamamoto were running the meeting. Sue was the chancellor at UCSF at the time. Keith was running strategy and vice chancellor for research. That's the day they said they started to recruit me, was four years.

[00:21:29] But it was a lateral move to me. Why leave? Stanford had literally everything at Stanford. Why move to UCSF? Until I was introduced to a man named Jack Stobo, and Jack Stobo was over at UC Oakland. So that's where the office of the president was. And Jack Stobo was tasked with running subject called UC Health, the University of California Health System.

[00:21:51] Never heard of this thing before, but I got to meet him as part of my recruitment and he said, we're building this umbrella across the entire [00:22:00] University of California Health Enterprise. And at the time there were 18, now we have 20 health professional schools, six medical schools in that set. We have now 12 hospitals and growing.

[00:22:10] And Jack said you get to build a central data warehouse of all that clinical data for the entire University of California. That job did not exist anywhere else, and I'm still gonna argue does not exist anywhere else. And moreover, this job will never work in Boston or New York City 'cause you all still compete with each other in your hospitals and we don't in the University of California.

[00:22:35] And I signed the offer letter one week later. That was it. And so that was the December of 2014. I started at University of California on April Fool's Day 2015. I've been here for about eight and a half years so far. Wow. So I, I can't argue, um, with you that that seems to be like a uniquely UC type of job description.

[00:22:59] Sadly, [00:23:00] I'm very convinced that you are correct in that. I think let's pause there though, 'cause I do wanna come back. Yes. And really dive into your role as chief data scientist. But I think before we do that, we wanna give folks a sense of your research agenda and sort of how you got to, to be the king data scientist in California.

[00:23:15] Why just California?

[00:23:20] I love it. So, Atul, before we just jump into your, your research and your lab's work, I have to say I've had the pleasure of seeing you speak on a number of occasions and you give these electrifying talks that are really, really exciting that I think make a, a lot of your audience want to just get up and go build something, or go do something, or go write a paper.

[00:23:40] And so did I hear correctly in that amazing story you just told about your recruitment to UC that it started with a keynote that you gave at the precision medicine meeting? That is correct. It was, I think, if I remember correctly, it was, goodness. I forget who else were the speakers.

[00:23:56] I could go find the details. I gave the first [00:24:00] one. And so just go slightly deeper because it's gonna tie into the research. At that time, at 2010, 2011, I was already at Stanford for five or six years and I was totally building my career around publicly available data, right? And at that time,

because of the genomics revolution, we still see some echoes, many louder echoes.

[00:24:21] Now people were making folks release their data to the public, especially in molecular world gen bank. We had something called the gene expression omnibus, where these RNA measurements were going. And at a certain point I realized that I could tell you how I got to this point because it's also a funny story.

[00:24:39] I realized I didn't need to go collaborate with folks. Of course, I love collaborating with folks, but they have to share their data with me, and I can do more things with their dataset than they might've thought possible when they wrote their first paper. And so we were gathering datasets like crazy, curating them, auto curating them, doing all sorts of magic with [00:25:00] the disease datasets and realized,

[00:25:02] you know, we could start to build a new taxonomy of human disease, given the molecular view of these datasets of these diseases. What does that mean? Right? So, this is how I was launching my talks back then. I literally, in my bag, was carrying these ancient books, a hundred, 150-year-old books of how we used to classify diseases, right?

[00:25:23] Based on symptoms. And, okay, the doctors call cardiologists take care of these diseases. Let's put them in one chapter. And the pulmonologist, they're in a different chapter, right? And it has nothing to do with the molecular side of things. So that's where I'd launched my career while I was at Stanford.

[00:25:39] We'll get to the drug side in a moment, but that is what caught Keith Yamamoto's eyes and Sue Desmond-Hellman, and Zak and many others. That's why I gave that talk. That's what led to me moving to UC four years later. And I, I just, that's amazing. I just wanna inter-, I just wanna interject and like reiterate something that Raj said and that your talks are electric.

[00:25:59] I was going [00:26:00] back and watching a lot of your talks on YouTube to get ready for this. And I had to stop because my heart rate was too high. Like, I was too jazzed about what you were saying. I could only do them in short bursts because they were so energizing. It's practice. Okay, so look, I, I did debate team in high school and I loved it.

[00:26:16] So I love public speaking. I still remember the first talk. Zak sent me to another thing. A good mentor does. I gotta do better that at that, no doubt. But Zak couldn't make some congressional testimony thing, and literally I

hadn't even joined his lab. He says, Zak, so Atul, go give this talk for me. I'm like, boy, oh boy.

[00:26:34] This is how it's gonna be. And so I learned, because again, genomics has taken off. Zak could only do so many of these. How many could I do? So I got a lot of practice. Practice makes perfect. A lot changed when I had to give my first TEDMED Talk in 2012. So I watched a hundred or two hundred TED Talks and TEDMED Talks.

[00:26:54] That's when I think my game really went up because I learned the style I liked. What is my [00:27:00] style? I pause during talks. You need people to think, you know, you want them to look up off their phones. And, uh, I love history in talks, uh, which is why I talk about how we classify diseases way back and always bring a prop to your talks.

[00:27:16] Uh, because when you're holding something up, people have to look up, you gotta get off their laptop. So I, I, I know this used to be a microarray, but in the age of AI, what is the prop? Is it a GPU? You're asking a great question. What do I bring for a prop for AI? I haven't figured that out yet. Uh, that is a great question.

[00:27:33] I still carry a microarray in my backpack. It's in the main pocket. So I think, um, uh, that's a great question. I, I need to think of an answer perhaps by the end of this podcast. I'll have an answer. But you have to have a prop. You have to have a prop because otherwise people are just heads down. Maybe they've been tweeting about you but their heads down.

[00:27:55] Great question. I'll think of an answer there. Alright, great. So, Atul, so first I just have to [00:28:00] note. That it's a great reminder to us, and I think a great note for all of our students, postdocs, other folks who are listening, I think, talks take practice and, you don't necessarily find your style and your preferred approach until, maybe a little bit later on and watching lots and lots of other talks.

[00:28:20] And so take every opportunity to practice giving talks, give them regularly, flex that muscle, exercise that muscle very regularly. And I also have to say, I, I liked hearing that this sort of secret, I think you just, uh, disclosed here, which is that you watched a hundred TED Talks, figured out what your style was, and then emulated that, and then practiced that.

[00:28:39] And I think really have honed it and just look at the outcomes, right? These are not just lines on your CV. These are transformational, arguably career

switches and changes in new opportunities that come from being able to give wonderful talks. So, I also have to say, I think this is a great, great point to maybe transition into some of your research.

[00:28:57] You've done this amazing both breadth and [00:29:00] depth of work, from genomics to gene expression, to drug repurposing, to big data, AI, and medicine. It's impossible for me to pick one of your papers to start with, but I am gonna try and I'm going to maybe just pick this first paper as a branching off point and then let you kind of improv and throw some other papers into the mix here.

[00:29:21] So you've published a lot of really interesting, thought-provoking work and bioinformatics and genomics and big data and AI. One that I really like is this paper by Marina Sirota. Your former student. Now I believe you're a faculty member at UCSF. And this is a paper a little over a decade ago in *Science Translational Medicine*.

[00:29:39] Discovery and preclinical validation of drug implications using compendia of public gene expression data. And, so, to me, this really typifies the quintessential Atul approach to this problem, which I think is a very, very large and interesting problem. And you motivated a lot of other folks now to take similar approaches.

[00:29:58] Maybe could you first start us off [00:30:00] by just walking us through how you got started on this particular project, what the genesis of the idea was, how it came together, and as maybe also a transition to, you know, what this has led to outside of academia at work by your lab and your involvement in, in some companies that are furthering this along.

[00:30:18] Yeah, boy, I mean, we could spend a day talking about this paper and like the before and the after on this one. So first of all, you'll notice I love this paper. Uh, it's actually a companion paper. So we had two papers out at the same time. Marina is the first author on one of them, and Joel Dudley's, the first author of the second one.

[00:30:37] And they're both co-second on the other. So both came out at the same time. *Science Translational Medicine*, which was at the time a brand-new journal. I mean, it had just started to come out and we realized we could do this. So, okay, so how do I even start with this paper? Look, if you have the time, I will go way back to even before I joined Stanford, because what happened during my team, we have the time.

[00:30:58] We have time, we have the time, we have the time, right? We have the time. [00:31:00] So my very second paper, I'm writing with Zak on genomics. Zak was already saying, what are you gonna do? Just write another methods paper. And Zak right away said, you know, I had this first paper called Relevance Networks and we were like making correlational graphs and okay, that's great.

[00:31:19] And we had some findings. But Zak, right away planted the seed in me that you can't just keep writing methods, you're, you're gonna wanna use them and come up with the findings. And that was part of the biologist in me as well, right? I mean, remember I took that year off. I wanted to do some biology and some medicine, so I'm not gonna be a methods-only person, right?

[00:31:38] You can go through my whole literature. And maybe we have 10, 20 papers max on a URL. Like here's a website with something, something, something. Right. Most of my papers, especially ones you're making the list on here, are things where we discovered something using a method, right? A computational method.

[00:31:57] And not just that we're using data to do it, someone else's [00:32:00] data to do it. What happened is as I'm interviewing for Stanford, as I'm leaving Harvard, part of the counteroffer was coming from the Joslin Diabetes Center. I still remember, I'm having lunch with one of the faculty there, and I was talking about how I was doing this, right?

[00:32:15] I, it's funny, I was working in Ron Khan's lab and Zak's lab and everyone's lab, you know how it works there. And I was working on his dataset, one of his mice datasets on diabetes. I literally had to leave and cross Longwood Ave to go to the Beth Israel Deaconess Hospital to talk to another lab working on their diabetes dataset.

[00:32:33] They had the human dataset there, right? And uh, it turns out it was Jeff Fly who became the Dean of Harvard Medical School later, right? And so I realized I was the only one that had both people's datasets, right? I had the one and the other. And look, these two folks don't hate each other, but they're not, not necessarily gonna collaborate either.

[00:32:52] I was the only one with both datasets. And so what I realized is if you are the one with both, you can work with one, work with the [00:33:00] other. But then when you have both, you could write a story that has both people's datasets and you're not even stepping on people's toes because they get another paper with one dataset.

[00:33:09] No one says no to that, especially at Harvard. And then you could say, and then when I was having lunch and interviewing with the Joslin Diabetes Center faculty member. And I still remember the, the moment he says, you don't even need to talk to either of those two folks because everyone has to share their data.

[00:33:27] You could just start with the public data and then talk to them when you need something validated. Right? And so that's how I started my initial emails. I started this in the Harvard ecosystem. I moved to Stanford ended it this way. I would say people needed help analyzing data. They always do. I'll help you.

[00:33:45] My folks will be middle authors, maybe a start, second author if we merit that. But any other paper we write, we are gonna be the first and senior authors and you're gonna have to help us validate our findings. Right? And so [00:34:00] that's how the mantra started. And boom, that's how we got to these two papers here, because what we realized, we had so much disease data out there on the internet, so much drug data out there on the internet.

[00:34:10] We're gonna put these two together. And that's where Marina started this, as a rotations student. So that paper was a rotation project for her that ended up becoming her graduate project. And then obviously she's a faculty member now. And by the way, another mentor kind of lesson out there: every project is useful.

[00:34:30] Even rotation projects, we do not give out busy work to anyone. If you bust your butt, you're gonna get a paper out of a rotation project in my lab, too. We don't want busy work. Anyway, we had so much data. We started to come up with predictions that drugs, including old drugs, could work for new diseases.

[00:34:47] And then we realized that we just didn't wanna make predictions. We're gonna have to test them. And so we had collaborators to help us, right on the GI side, uh, that was the one paper for inflammatory bowel disease. And then we had core [00:35:00] facilities on the cancer side that we could use to test out our cancer drugs.

[00:35:04] Another lesson learned, right? If the finish line in our field is too easy, push out the finish line, right? Right. If everyone is starting to make predictions of things, then start to actually test them, and test them in mouse models, and then you'll see in subsequent papers. Not just test for mouse models, test them in clinical trials, right?

[00:35:25] ‘Cause the finish line is getting easier and easier. If you don’t push the finish line out, someone else is gonna push it out for you. Right? And so that, again, lessons learned that we wanted to keep pushing the story out as far as we could. I think we got two back to back-to-back *Science Translational Medicine* papers.

[00:35:43] Because we had two predictions using public data, and we tested those predictions in animal models. And so those two papers, I think they’re close to being cited a thousand times each now at this point are, are certainly over 500. I think those are gonna go down as really [00:36:00] some, some of the best papers that came outta my lab because we really, we started with other people’s data.

[00:36:04] We realized we could empower those datasets and those researchers now and really test these things. I still think we have too many bioinformaticians who are just, are not comfortable testing things themselves. It turns out you can hire core facilities, you can hire contract research organizations to do this work.

[00:36:23] I think I had a TEDMED Talk on this one. Or you can find collaborators to get this work done. Why are you just making predictions? You gotta test these things, too. I just, I wanna kinda like blow up what you just said there so that folks have an appreciation of something that you kind of just like alluded to.

[00:36:38] I believe you have a previous life as a market maker almost. Or the equivalent of like, I need lab services, uh, where you can go and drag and drop an assay into a shopping cart and hit checkout and some CRO will, they’ll do this. I think that’s what you just mentioned in passing, but that in and of itself is an amazing, uh, would be someone else’s entire career.

[00:36:56] Yeah. It turned out this company called Assay Depot came out at the [00:37:00] time again, having smart people in your lab. I think Joel or Marina found this website, realized this could happen. Assay Depot is now known as scientists.com. There were others that were starting to come out at the time in Palo Alto where there were two-sided marketplaces for contract research organizations from core facilities or from other labs.

[00:37:20] Yeah. And all of a sudden we realized we could get this kind of work done, but this is how you had to do jiu jitsu. Okay? This is the jiu jitsu you have to do as a faculty member. Right? Very quickly people could say, oh my

goodness, what kind of quality research you're gonna get from doing this kind of research in a contract facility.

[00:37:39] We have experts that are doing this in our academic labs. The jiu jitsu is, I can hire 10 of these labs to do this work. Right? You can't do that. I can, and then I can figure out what's in common here, right? I don't have to believe it. If it's just one lab, I can hire five labs, 10 labs, right? If it's only $10,000 or $20,000 to get this experiment done.

[00:37:59] So [00:38:00] you have to know how they're gonna criticize you. And it's not like you have to fire the first shot. I just know you're gonna try to poke on that and I'm just gonna poke right back ahead of time. So yeah, if it's a contract research organization, no one says you have to hire just one of them, right?

[00:38:16] And so you have to understand ahead of time where the criticism could come from and just head it off, you know? Uh, a good friend of mine, Carlos Bustamante, uses his phrase "iron fist in a velvet glove." So I love that phrase. That's, that's kind of me too. Iron fist in a velvet glove. Maybe that's too harsh. I love it.

[00:38:38] I love it. Um, so it's a tool. You've published these high impact influential papers, but then they've also led to the creation of companies that you've helped start. And so I think one of these is NuMedii. Yes. That does work on drug repurposing. There's a host of others. I think you have a COI slide that you start off with that you're proud [00:39:00] of, that's inspired by George Church.

[00:39:02] Right. And, and George Church's opening to many of his talks. So could you maybe describe how you think about problems as a problem for your academic lab versus when it's hit the point where you need to start a company and this is something that has to be taken over by a commercial entity as opposed to, to your lab?

[00:39:20] Yeah. Again, it's academic jiu jitsu, right. So, you know, and that's one of the best things about being on this coast compared to Boston. And again, I, I love Boston. I love the ecosystem there. But in some ways, and I learned this really kind of watching my mentors, it's almost taboo to wanna talk about companies in Boston.

[00:39:39] I know people start companies in Boston. I get all that right, and the whole beltway system there too. But in the Bay Area, first of all, it's okay to

talk about it. In fact, it's encouraged and it's beyond that, right? You only have one company, you're a loser, right? I have two. Here's how I do it, kind of thing, right?

[00:39:56] Especially in Stanford and UC, Berkeley, [00:40:00] UCSF, right? So, first of all, it's, it's, it's encouraged much more out here. And that appeals to me. And then second of all, just to talk about that conflict-of-interest slide. Again, academic jiu jitsu, right? People are ashamed of that slide, right? So I take that as a point of pride now, right?

[00:40:16] That's, that's jiu jitsu here, right? I'm, I'm going to use that. Uh, and you know, again, I say, here's my point. Here are my conflicts of interest. And I have just a few. And I pause 'cause everyone looks up from their phone and they start to giggle, right? And all of a sudden they realize, right. So all that slide is meant for faculty to tell people I'm not trustworthy, right?

[00:40:37] That's why you put that slide of shame up there, right? I'm not trustworthy and I'm gonna, again, jiu jitsu it out there to say, look how many companies trust me, right? So fine. Don't believe me. I literally say in every talk, I don't blame you if you don't believe another word I say over the next 45 minutes or so, right?

[00:40:56] I literally say that, that phrase, but all of these [00:41:00] folks trust me. And you know what? Some of them hate each other, but they still trust me, right? So this is how that's jiu jitsu, right? You know, people are gonna kind of poke at you on this. And in some weird way, again, I don't know how academia works sometimes, it is more trustworthy to have 800 companies on that slide than one company on that slide, right?

[00:41:19] It's such a weird thing. Okay, so how does this work? Right? And I also say this on that slide, you've heard the talks I tell my lab, if you wanna change the world, you can't just keep writing papers about it, right? If you've invented something, you have to file the disclosure. I try to get all my grad students to file at least one disclosure or my postdoc so they learn the process.

[00:41:38] And if no one licenses it, then it's gonna be up to you to start the company, right? Because otherwise, why do we do this work? Now, this is a complicated environment here. Complicated story. There are certain rules you have to follow. At the same time though, we're allowed to be applied in medical schools. I will say it is much harder to do this in a College [00:42:00] of Arts and Sciences or any of these other kind of schools within our university.

[00:42:05] But in medical schools, right? We know what patients have to go through in an exam room every single day. It is okay to be applied to make this better for patients and families and doctors, right? So in a medical school, we're allowed to do this, but there are these standard six rules that you have to follow, and I could go look them up.

[00:42:24] Like you can't get your grad students to start a company while you're in that company. You can't do clinical trials on your own patients. You have to be transparent with everything. You have to file all the forms, right? There are these standard rules. It turns out that the same basic six rules, whether you're at Harvard, Stanford, or the University of California, okay, they're not that different.

[00:42:45] There are minor differences here and there, which I know the letters of all the laws here and all the rules, but in general, we are encouraged to do this, to spend our 20% of our time consulting or really starting companies. So [00:43:00] I'm a believer in entrepreneurship. It goes way back to my summer intern days at Apple, where they gave us for free this book by John Scully, who was the CEO at the time.

[00:43:10] This was post-Steve Jobs and pre-Steve Jobs, but it was about how John Scully was recruited to Apple and his whole journey, and all of a sudden I got enamored by business. I started my first company in medical school writing software. Sold a whole bunch of it, helped me buy a condo in Jamaica Plain in Boston.

[00:43:29] I was never gonna IPO or anything like that. I was writing software for Boston Children's on the side, making some side money there. So I was into entrepreneurship even in medical school. And so of course I'm gonna communicate that to my mentees that we can do this and they should do this. I don't make anyone do this, but I think we tend to attract people that wanna do it.

[00:43:49] You mentioned George Church, but there's many others in our field, but actually, I'm not gonna say many of us, too few others, actually. Bob Langers, another one. We have, uh, [00:44:00] Rick Myers here over in the Bay Area, Stanford. And then of course we have such history of companies that have been started in the Bay Area.

[00:44:08] I'm literally right next door to a building called Genentech Hall, and Genentech was our startup company at UCSF, of course, a small one at the time, I suppose. Lawsuits nonetheless, but still, I guess we're still friends, uh,

because we named a building. And so you see the history here, you see the role models here.

[00:44:28] If you choose to want to become like that, then it's really kind of easier to do, I think, on the West Coast. And look, it can happen in Boston for sure, but it seems more hierarchical there, uh, than here. And we could talk about those differences if those are of interest. But let me just pause there and see where we're going next.

[00:44:46] Yeah, no, I, I think that was a, a good overview of I, because I think that that pipeline from academia to startup, is opaque to lots of people. And so I think you did a beautiful job there of spelling out the motivation, how it [00:45:00] actually kind of works, when an idea may be start up-able, I guess is one way to describe it.

[00:45:05] Um, and then just like I think I, again, I think de-stigmatizing it is the right is, is kind of an important thing to do. And I think that you have more than anyone else I know been very good at de-stigmatizing entrepreneurial tendencies in academia. Yeah, I tell, I totally try to do that. I, I say this all the time, it should not be taboo to talk about companies on stage here, right?

[00:45:25] You have to follow all the rules and you have, you know, all, all of that. You have to follow the rules, but should not be taboo. And we should be encouraged to do that and we should encourage our mentees to do it. Now at the same time, not everyone can do this. Not everyone has the job in which you have the privilege to start companies, right?

[00:45:40] If you're a clinician, you're spending, uh, pajama time writing your notes at home for your patients. Who has that, who then has the time to do this, right? If, uh, little kids at home, it makes it much harder, right? So it takes time in life to get to the point where you can do this easily, you can do this more easily with your trainees.

[00:45:58] And so it's not the [00:46:00] case. Again, it takes privilege to be able to start companies and, uh, more privilege and more luck to get them to be successful. But it should be something we do. I can bring up another story too. Uh, why, why, another reason why I think we should be thinking more about companies, especially for grad students, have you ever thought about why on earth does the federal government give us money as grad students to go through graduate training?

[00:46:25] Right? What's in it for the government? Why does the federal government give out training grants or F awards or training, any of these things? Why does the federal government, presumably one half of the country, I'm sure would have one answer, but presumably it's because there's some expected ROI on having a highly skilled scientific workforce and having economic returns on the basis of that.

[00:46:48] Absolutely. Well let you know. So I had a really good friend in India who told me the story and gave me his analogy, his story from India, and he went to IIT Mumbai, which [00:47:00] is incredibly impossibly hard to get into, like less than 1% acceptance rate. But once you're accepted, especially as a grad student, you're fully funded there.

[00:47:08] And, uh, he told me the story, how they were introduced to this concept in India, and he was brought in on the first day to the big auditorium. And they're surrounded by all the other students, all the other grad students. And he was told that chancellor was there, I suppose. And he said, he said, look around, you're the creme de la creme, the, the cream of the crop in India.

[00:47:28] You are all here and the government pays for you to be a student here at IIT. Why does the government pay you to be a student at IIT? And what did the chancellor say? The government isn't funding you to go get a career when you're done at IIT. The government is funding you to create the careers when you leave

[00:47:54] IIT. And I was thinking in this story, boy oh boy, how we [00:48:00] would be very different if we thought that way. Because right now I still think, again, there's no mean intent in the sentence I'm about to say, but I think we think in a very selfish way when we get a training grant, and I've been on many training grants that we keep thinking it's about us and how do we build our career and how do I get those papers out and those annual reports.

[00:48:22] But boy, I really actually think, why isn't it the case that we also teach people the awardees of these career development awards and training grants, that they should be creating other careers for the United States. You know, when the government says we need to create more jobs, we are how they should do it, guys, just because we're in bioinformatics and labs, that is us too.

[00:48:43] So believe me, I'm a believer in all of this. I, this is what I communicate. This is not just a, an also ran, right. If we've invented something and we wanna help people, we have got to get it out of our labs and every rule

lets us do this in a safe, responsible [00:49:00] way. So, yeah, uh, it's not just a, an also also ran here.

[00:49:04] I think this is something that should be part of how we write a paper. I think about it that seriously, I, I mean, I agree not just from a, I'm totally with you. I want to not just write papers, but make translational impact on real people's lives. But I guess too, I've never really thought about how empowered you are from an academic perspective, from a training grants, but also from like Dole Bailey and IP ownership and things like that.

[00:49:27] It actually is like a really, I think, unique and privileged position to be able to be like a real economic driver. Yeah. Okay. So first it's the Bayh-Dole Act. Uh, that's it. I, we'll fix it in post. You'll fix it in post. So first of all, think about, okay, so think about your institution. Think about my institution.

[00:49:46] We have these technology transfer offices, right? And they help file the disclosures and some become patents. And you ever wondered why these universities and hospitals have this, right? And usually when you ask folks, you think, oh, they wanna make [00:50:00] money. It's 'cause it's federal law. We have to have these offices, right?

[00:50:03] That this is, now again, people can poke holes at how the legislation has been written. The public can be cynical here. But before Bayh, right? Every, uh, faculty was just writing their own papers and everything was in the public domain, and no one could commercialize anything. But in meantime, if you think, again, you could be cynical, I choose not to be, that $1 from NIH goes so much further.

[00:50:27] When it helps your career and it gets a paper out, and it gets an invention done and it launches a company, and that company launches a hundred people right in their careers. That $1 from NIH went so much further. Because of the Bayh-Dole Act and, uh, the subsequent, uh, kind of practices around it. Now, again, you can be cynical.

[00:50:45] There are many cynical people out there about this process. I choose not to be. Alright, so I, I do wanna touch a little bit on picking back up the thread of your role as the chief data scientist for the UC system. Could you tell us just a little bit about what that [00:51:00] job role entails? Like what are your actual responsibilities?

[00:51:03] What are you, what are you trying to accomplish, um, in that role? Yeah. Okay. So first of all, I'm still a faculty member here at UC and UCSF. So

I start to write papers and grants. Uh, and I have a lab and I have grad students and postdocs. I also run my institute at UCSF. So we have about 90 faculty, all the data compute folks, recruit them, retain them, build out our infrastructure at UCSF.

[00:51:25] And then I got this role across the whole UC, which is what really convinced me to move to UC. And so what is the chief data scientist? Remember back in 2015, 2016, 2017, we started to realize we could pool all of our data in one place. Initially, we were pulling that data for operational and quality improvement reasons, which means how do we save money in delivering care?

[00:51:48] Like what, where are we wasting money? Why are we ordering this drug when this other drug could work? So it helped for us to get all that data in one place to start to ask and answer and fix some of those [00:52:00] problems. We start to, when you deliver care to folks, you have to report on the quality of that care a lot.

[00:52:06] Medicare, Medi-Cal, which is our Medicaid system here, and we can do that centrally, or at least work centrally to get that work done in a more efficient way. So we needed that data in one place to start to get, be more efficient in how we measure the quality of care, how we fix that quality of care, and how we report it.

[00:52:24] Then we started to open up for researchers, and so now hundreds of researchers have access to that data. It's all in the cloud. They can sign on to that using their own campus credentials, and we could talk about what that infrastructure looks like. But really it started by getting all that data in one place.

[00:52:40] Every two to four weeks, we're moving all that data centrally from each campus's Epic instance, some share Epic instances. We harmonize that data in collaboration with them, and then we have that central data warehouse at last count, it's just approaching 9 million patients that we've treated over the past 11 and a half years.

[00:52:59] I can go through some of [00:53:00] the more statistics. I'd have to pull them up here. Well, I think there's one statistic in particular I'd love to bring up, and it's, you do this amazing thing where you live tweet your keynote every time you give a keynote. And I think what I've noticed over time is the title of your talk is something like, uh, uh, trillions.

[00:53:15] Trillions, yeah. Trillion. So it, it's gone from millions to billions to I think now trillions. Can you tell us a little bit about. What goes into 1 trillion measurement dataset, and then all of the things that that captures about the patient experience. First of all, that title is completely made up, right?

[00:53:31] I love alliteration. Okay? So like, I think I've put in trillions and like other T words in there. So I like those kinds of alliterative things in how I give talks. It's another thing I learned from, uh, watching all those talks, but, uh, you know, it depends how you count these things, right? If you count every bite of every dicom file, every radiology file we have, it's probably in the terabytes and trillions, whatever.

[00:53:51] So it's a made-up number anyway, but we do have a ton of data in UC and across the University of California at our disposal. Now, first of all, I'm always gonna use the same [00:54:00] phrase. You'll probably hear this in many talks, safely, responsibly, respectfully using that data, okay? If we can't get this done safely, responsibly, respectfully, everything falls apart.

[00:54:09] So that is first and foremost, it's gotta be done in a safe way. Cybersecurity threats are crazy right now. Uh, and we gotta do this in a respectful way to make sure we have patients actually watching what we do. I have my own oversight board. There are patient representatives on there, ethics representatives on there.

[00:54:25] So we wanna make sure that we are doing this right by patients. At the same time, though, we are now in the kind of mantra now where if we don't do this, something's wrong with us because if we don't use this data to prove the practice of medicine, what a tragedy it will be given the billions we've spent on electronic health record systems.

[00:54:42] So that's the narrative I tend to use in my talks. Again, safely, responsibly, respectfully, but I say this all the time, this is most expensive data in America. We're literally paying doctors in their pajamas to write these notes. It will be a tragedy if we don't use this data to improve the practice of medicine.

[00:54:59] That [00:55:00] that's, that's where we start from, and then we get to all the kind of use cases from there. So I think that's a perfect, um, a perfect way to think about it. And I think also I'd like to, a lot of researchers we speak to are working on their own dataset or working in a single institution, but you have this unique perch across this vast health care system.

[00:55:19] How do those principles that you just outlined, how are they gonna impact AI deployment at a health care system level? So like, how do you think about rolling out what's happening with AI across a health care system like you see, are there any unique challenges there or like how are you thinking about the opportunity that AI presents at the sort of health care system level?

[00:55:38] Okay, so very complicated question. So first of all, I think it's safe to say we've already made an impact across the whole UC health system. We have an annual report and I'm gonna point folks to it. If there's any room for URLs in your kind of liner notes, uh, I'll give folks URLs so they can look to see, because we document and share with the public all of our use cases with clinical data.

[00:55:59] That [00:56:00] gives other hospital systems ideas what they should do with their data as well. So now AI is a new big thing, right? It's so old, it's new again, like everything else that in our field. And I think you can have little AI or big AI, right? So little AI people have been building models for a long time.

[00:56:17] We certainly have 400 researchers in the University of California system that we can point to that are interested in AI in biomedicine in particular. A couple hundred of them use our datasets, our central data, but there's other molecular datasets and folks out there. So AI has been around and it's certainly coming back.

[00:56:34] But I think there's this whole large language model aspect of it, which is definitely a new direction that I think we're, uh, trying to pioneer in. Now we have a couple advantages at the University of California, uh, and certainly because I have this role at UCSF and UC wide, we could kind of develop things at UCSF with the resources we have there and kind of smear it to UC wide, if you know what I mean.

[00:56:59] [00:57:00] Uh, to use, let's say, uh, schmearing lots. I heard that applied to cream cheese before, but never to LLMs. So never to LLMs or data. Right. Well, we can smear what we learned at UCSF. Now look, uh, one of great part parts of being at UC, uh, proper, let's say UC, op the office of the president, is I get to pick and choose the talent we get to hire.

[00:57:18] So actually most of my team building this central data warehouse is at UC Irvine. They're not at UCSF because they were way ahead with OMOP, which is OMOP, how we save all our data. It's a, a vendor neutral standard that

people use and we love it. But we've made a lot of extensions on OMOP. So that talent was at the University of California Irvine in their health system.

[00:57:40] So we tap into that. But at UCSF, we have a particular talent at de-identifying our clinical notes, and that's a killer advantage right now. So we started in 2016 just literally de-identifying all of our notes, and now we have 140 million of [00:58:00] our clinical notes legally de-identified, externally certified as de-identified.

[00:58:06] I know many others are trying to get to that point. We are already there. So we can actually do all sorts of magical things safely, responsibly, respectfully with our notes, with LLMs locally or with certain parameters that we set with OpenAI. They can't save anything. We're not transferring any of these notes to them.

[00:58:27] But we can use prompts in a safe, responsible way with OpenAI through Microsoft Azure. You can believe we're doing a lot of that right now. Um, so just a couple quick questions there. So that was gonna be my follow-up question, is that it sounds like your de-identification procedure is so good that you're okay sending it over the wire to OpenAI servers and letting their LM's chew on it and then give you back the answer.

[00:58:50] It sounds like the answer to that is yes. Uh, do you have any compelling demonstrations given the volume of notes that you could share given the head start that you have in this area. [00:59:00] So, okay, so first of all, to be really precise because people will hold me to these things. We have our data here at UCSF there, it's behind a firewall, even though it's de-identified.

[00:59:10] And just to be also very clear, our data at UCSF is also de-identified. Our IRB states it's non-human subjects research, so no further IRB even needed for further research with this data. Okay, so that's the level we've gotten. We have an external company called Archer Hall, which is certified this as de-identified.

[00:59:29] And believe me, when they were certifying, they would ask for a million notes at a time from us to look through some to them for identifiers. Okay? So again, we've been doing this since 2016. So what does it mean for us to use OpenAI? Well, we send a note or a query with a note to the OpenAI servers, GPT and

[00:59:49] believe me, I did not say ChatGPT. This is GPTAPI. Through the Azure servers and you have to have two flags turned off. I'm not gonna go into those details that tell them and make [01:00:00] sure they don't save it. Okay? When those two flags are off, they send you an answer back and they don't save anything.

[01:00:06] We save our answers and we move on to the next question. Now, what are some of the use cases here? Some of which have been already submitted to your journal, some of which have been submitted to certainly the top tier journals are looking for these papers. We have submitted some to yours and we will have submitted some to ours.

[01:00:23] Mine, I though they, I don't really own this journal. I'm just a humble editor. What are some of the use cases? In my lab, we are looking at side effects of drugs. Why do patients stop a drug? Right? These are things that a human could learn. It would take a lot of humans to curate that information from a database, let's say text notes and LLMs can find some of that really amazingly fast.

[01:00:46] What are some of the biomarkers in this cancer patient's progress note? Can you actually triage a patient from just the history and past medical history and review systems in an emergency room setting? Well, let's try 10,000 of [01:01:00] these, right? So we can do all sorts of these now in a safe, responsible way,

[01:01:04] 'cause we have those notes. Now that's UCSF. We're trying to get the notes work in UC wide. There's pathology notes, there's radiology notes. Can you put the image with the note? Obviously everyone's trying to do that right now. We can do that, too, because our images at UCSF are also legally de-identified externally certified non-human subjects research.

[01:01:27] So we've been starting this for a while. We kind of knew where this field was going back in 2016. This is why, this is why we've done this. By the way, our cancer genomes are also done the same way. So we can put all three of those together, uh, for some interesting stories. So I suspect Harvard hospitals are working on these.

[01:01:44] I get the emails from the people looking for help, and maybe you'll get there. I, I'll say this is, this has now become implicitly a very compelling advertisement for moving to the West Coast and working in the UC system. So, yeah, and just to be clear, so again, UCSF, we go deeper. [01:02:00] Right now

we can, because we have the notes, we have the images, we have the cancer genomes, uh, de- and structured data UC wide.

[01:02:06] We have just the structured data de-identified, again, fully de-identified and non-human subjects research. So any grad student on any of our health campuses, and if you're not on a health campus, you get the health credentials by going through training. Even a grad student can access this data, single sign-on it's cloud-based Databricks, which is kind of like a Jupyter Notebook and no IRB needed so ready to go.

[01:02:29] But if you're outside UC, you cannot have access right now. So yeah, it's an extremely compelling reason to join the UC on any health campus by design. By design. So Atul, we have a little bit of a surprise and this is the lightning round. So are you ready for a rapid fire set of questions?

[01:02:55] The only rule is that these are lightning round questions, so short responses, one [01:03:00] to two sentences. They are a little bit all over the place. But I think should be pretty interesting. Okay. And I'm sure if I need to pause to think of an answer, that will come out in the post as well. Perfect.

[01:03:13] Alright. The first question, if you weren't in medicine, what job would you be doing? I'd probably be in some kind of software engineering, uh, of some sort, or I'd probably join a company that I started. So health, health IT still incredible time. Lots of companies left to start. I'd probably be in one of those.

[01:03:32] Alright, I'm not sure if you're familiar with the biblical story of the flood of Noah, but the way this story ends is that a dove returns with an olive branch in its mouth to let the people know that life is returning to normal. You are a prolific traveler and the dove from the flood of Noah story for me was when you started tweeting about United again.

[01:03:49] You usually live, tweet your travels. So what I'm wondering is if you have a set of tips that can make traveling less painful, because I do not like traveling. I do not like [01:04:00] flying, but you seem to do it so well. So you must have some tips to pass on. I have a lot of travel tips. So first of all, I travel a lot.

[01:04:06] I used for a couple years. I used to be Global Services on United. Now I'm a humble one k uh, party. You'll still make it this year. So I travel a lot. I tend to love aisle seats. I tend to get to airports early. I have a United Club

membership. Look, the way I judge an airline isn't when things are going well, it's when things are going poorly.

[01:04:28] Try to get all the help you can by going to the Club if you can afford it, or any of these kinds of folks where that's where they help you the most. I've learned this from others. Find your favorite restaurants in the airports who travel through the most. I love Rick Bayless at O'Hare, even though O'Hare is a nightmare to connect through, I'd rather have the food there when I get stuck compared to Denver, which has like no good food.

[01:04:51] What else can I tell you? Um, let me pause. Let me think. I think that in itself, yeah, that already was more considerations. I think that measuring an [01:05:00] airline by their worst case performance is also like good information. Uh, yeah, exactly. It's irregular operations, uh, that you kind of judge an airline by.

[01:05:09] Yeah. Shy alternative airports, West Coast folks constantly forget that we can fly directly to national, not just O'Hare, not just Dulles. Um, I could go on and on and on. How could that be a rapid round question? Fair. Fair enough. Atul, will AI and medicine be driven more by computer scientists or by clinicians?

[01:05:32] It's gonna be much more driven by clinicians who know enough about informatics to start more companies to get the use cases done. Definitely not computer scientists unless they choose to figure out these use cases. In fact, I think that's row number one. I have this whole complicated slide of advice I give to AI scientists.

[01:05:52] I can put that in the liner notes. I'll tweet it out again. But always find two doctors that validate your use cases. If you happen to be a computer [01:06:00] scientist in this field, don't just go with your uncle who says this is a useful use case. Find someone else too, 'cause otherwise we keep getting the same use cases again and again.

[01:06:09] So, corollary of your assay depot, you can run it more than one time. Yes. You need two or three docs to really validate a use case from the computer science side. If you're a clinician side, I don't think they think enough about how AI is gonna change their world. Uh, I think that's the problem. I think it's gonna change it a lot more than people know.

[01:06:27] So I think getting those clinicians, clinician, uh, investigators, and certainly the informatics field, we've gotta get them to start more companies. I

think that that's a pain point right now. It, again, we have too many people in our field just writing papers. I think we have some good candidates for the answer to this next question, but what is the best piece of advice you've ever received?

[01:06:49] Oh boy. Okay. Now let me think. I don't know. Okay. I could tell you the best thing I ever did is to marry my wife. [01:07:00] Okay? For sure. I married a, an incredibly brilliant woman, uh, Gini Deshpande, who is the founder of NuMedii, which is why I love that company the most, uh, with the drugs, drug repositioning, AI and drug discovery.

[01:07:14] You'll never be bored if you marry an incredible spouse and, and technically not an answer to the question, but we will allow it because it was such a good answer. So we will, we will allow it. Excellent, Atul. I love it. I get to ask this question. If you could be an Olympic athlete, in which sport would you compete?

[01:07:32] Oh, Jeez Louise. Okay, let me think. So first of all, I am terrible at sports. I don't even follow sports. I definitely don't follow the Olympics. Uh, but probably if I could work at it, I would probably do something like swimming. Uh, it's a nice sport. Uh, it's something you could do for the rest of your life. I would probably work at that the best.

[01:07:59] Um, [01:08:00] I am not, I don't have a good answer for you for this, this one. Oh, we'll, we'll accept it. Okay. Yeah. Okay. Last one. What did working as a software engineer at Apple specifically teach you about medicine, AI, or what lessons do you carry forward from your, your time there? That's a great question. It's okay.

[01:08:21] Slightly longer than rapid here, but it was a tough time at Apple. They had just gotten rid of the beer bashes. You can imagine the protests. Okay. Because they were losing money. Okay? This is the John Scully era. You know, they were trying to figure out what max to sell. This is even before Newton's, which I loved, and they didn't take off, sadly.

[01:08:41] Okay? What if was the lesson I learned? Leadership makes a huge bit of difference in, in organizations. When you're a junior faculty and instructor, assistant professor, you kind of don't really care who the dean is or who the chancellor is. You're head's down, you're doing your thing, you're building out your own story.

[01:08:59] But once you're [01:09:00] past that, leadership is almost everything. And even great organizations can turn on a dime. And so how you steward that leadership is so critical. But you don't see that when you're junior and when you're young. And yeah, it's easy to poke at people. You're making so much money as a

[01:09:17] CEO or as a board member, but those folks have incredibly important jobs.

[01:09:21] Now, will you say, is it worth all that money? You can argue that, but companies, institutions can change on a dime. It takes a lot to prevent badness from happening in institutions. And that I have to say, is leadership. And I learned that at Apple. That's a great answer. Alright, Atul, you passed the lightning round with flying colors.

[01:09:43] Uh, we have just a few last big picture questions. Sure. The first, I think you've really exemplified with your work, and you've already touched upon this several times, but maybe you can give us some, some last parting thoughts here. How has the democratization of data empowered patients? [01:10:00] How has the democratization of data empowered patients?

[01:10:02] So first of all, I think it's amazing right now that we give out data to patients at all. Okay. This is a change from when I was in training. And so you get one example of that in the Apple Health app, and I talk about this all the time, that you can go to Apple Health app. It's if you have an iPhone, right?

[01:10:20] Uh, it's the one that looks like a big heart and in there is medical records and you can go download your data. Using the same user ID and password from your hospital portal's credentials, and that's great, and it's given out in a federal standard and you get those lab test results. But if you go see many doctors, it's all in one nice timeline and Apple does that nicely for you.

[01:10:41] We realized at UCSF, just to put in a plug, that so many people did not have an iPhone, that we helped work with Common Health to get an Android version of that done. Right. So that's useful too. So I think it's amazing that we give out data to patients. Now we're giving out the notes to patients.

[01:10:58] It's not exactly in the Apple Health app, [01:11:00] but it's through the Epic or Cerner apps. And at the same time though, we're not explaining what these notes mean to people, right? Yeah. We do that when they come see us again in an encounter. So I think we've democratized clinical data going to patients. I, myself, use this like crazy.

[01:11:17] I have many friends who do this, too. But I think the future, I, maybe you've heard it here first or not, is decision support tools for patients. Okay. I know we work on decision support tools that we try to deploy into Epic and Cerner for docs. I think it's still absolutely a green field for apps and companies with patient-facing decision support tools built on top of the data we already give to these folks that we call patients.

[01:11:48] And by the way, it's federal law. We have to give this data to patients, right? So what's it gonna look like when patients get apps that help them interpret that data? Boy, that would be a great [01:12:00] use for AI and LLMs and all the rest. So I, I see the democratization just starting. Alright, and so this will I think be our last question.

[01:12:10] So Atul, obviously you've had a long and storied career, uh, you've had a huge impact in the field. I guess like looking forward, could you help us extrapolate over the next 5, 10, 20 years, how does the confluence of data medicine and AI play out? And what are you most excited about? Okay, so first of all, you're talking past tense.

[01:12:29] I got a long way to go here. Uh, so storied career. Well, I know, I'm just saying given your experience. Yeah. Given your experience, given the perch that you're at. Yes. Uh, synthesize for us and tell us what we should be excited about. So first of all, who can predict anything 10 to 20 years from now? Right. Uh, but certainly.

[01:12:47] Uh, we do have an, a very unusual health care system in the United States. I, I use this word system very generally here, right? We don't have a health care system. We have a whole bunch of two-way contracts signed by a whole bunch [01:13:00] of entities and the, uh, kind of meta properties that this is a health care system.

[01:13:05] So how much longer can this system last is not clear, right? And so what will take it over. I don't think we're gonna evolve to a better system. I do not think we can evolve to better system because all the entrenched participants, I'm not gonna poke fingers because I'm one of them. So our payers, so our pharma, so our device makers, so as everyone else, other trench players kind of do still love the current system, right?

[01:13:32] We do ask for more money every year from the ultimate payers, which are patients and small businesses, and the federal government. And somehow every year they keep paying the amount we ask for. And so I don't

think we're gonna evolve to a better system. It will take what they call a black swan event or something to change things radically.

[01:13:50] So if nothing changes in five to 10 to 20 years, we are gonna be in a health system that is going to cost more and more, we'll hire more and more [01:14:00] parts of the economy, right? Those people will be working for health care systems already. The health care system in many states is the largest employer in many of our states in the United States.

[01:14:11] And we will have more drugs and we will probably have more rules around using those drugs and devices because things will cost more and more. And the rules will be implemented by AI. So, you know, we will be sending prior auth letters in as physicians to an AI that's gonna adjudicate. But our prior auth letter will also come from our AI.

[01:14:31] So we will have AIs fighting AIs in this future system, which means more and more people hired more companies from whom we will purchase products, adding of course to the entire health care expense. Uh, so if nothing magical happens, I think that's where we will be in five to 10 to 20 years. But that's a safe bet, right?

[01:14:52] I think it's a safe bet. I think that's a safe bet. But do you, so Covid had the potential to be a black swan event that [01:15:00] was a structural shift that would cause change. Do you, do you think that there's any chance over the next, let's say 20 years that we have a black swan event or a structural change that gives us something new and maybe AI is at the center of that and maybe it's not?

[01:15:16] So first of all, even the first part of your premise, the question, I still am not sure how much Covid changed, uh, in our U.S. system. Well, so there was some hope that, that, that, that some people glass half full. We had a global pandemic, but at least maybe there'd be some change. Whether or not that's true I think is a different thing, but that is the type of black swan event that I, I think I was hearing you say something like that on that scale.

[01:15:36] No, I think more of a black swan event, uh, meaning political change in the United States, right? It's that so many people, so many, uh, frustrated folks, companies perhaps ask for a change and all of a sudden there's some political will to get it done. That just seems like an incredibly rare, hard to find black swan to happen in the United States.

[01:15:56] So, but that's, I think what it would take. Covid didn't [01:16:00] change the political landscape of the United States, right? So it would take that to change how things work here. In the meantime, the health care system's gonna get more and more expensive, more and more unaffordable, and there might be more and more innovations at the least expensive end, right?

[01:16:14] We will have more telemedicine visits, we will have more virtual first health care plans, right? You don't get to see a human unless you pass some virtual level. Then we'll have an AI first plan, right? I mean, so we'll keep having to innovate at the cheaper and cheaper end because the other side of it will get more and more costly.

[01:16:33] Those are easy predictions for five to 10 to 20 years. If nothing, uh, if nothing changes, I think I buy that. Atul, thank you so much. This has been a real pleasure and very enlightening. Thank you so much for being on AI Grand Rounds. Thanks for having me.