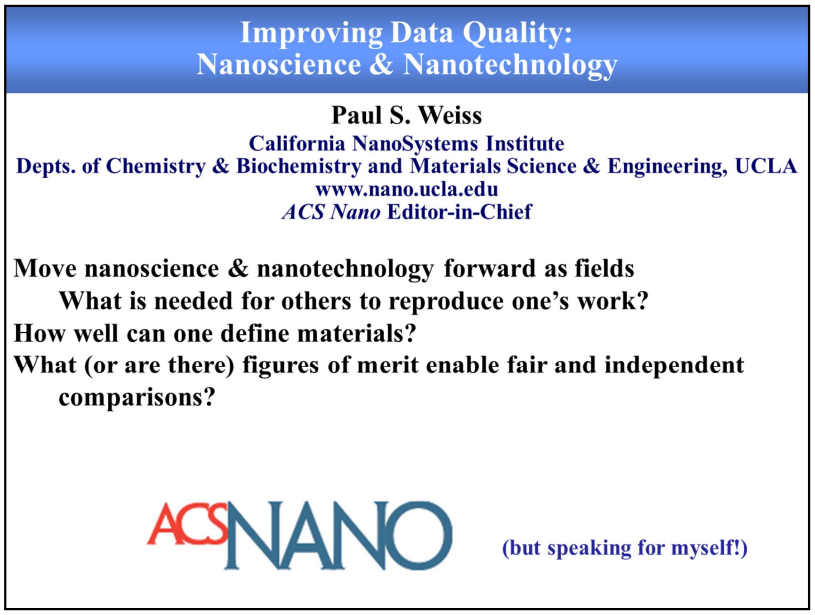
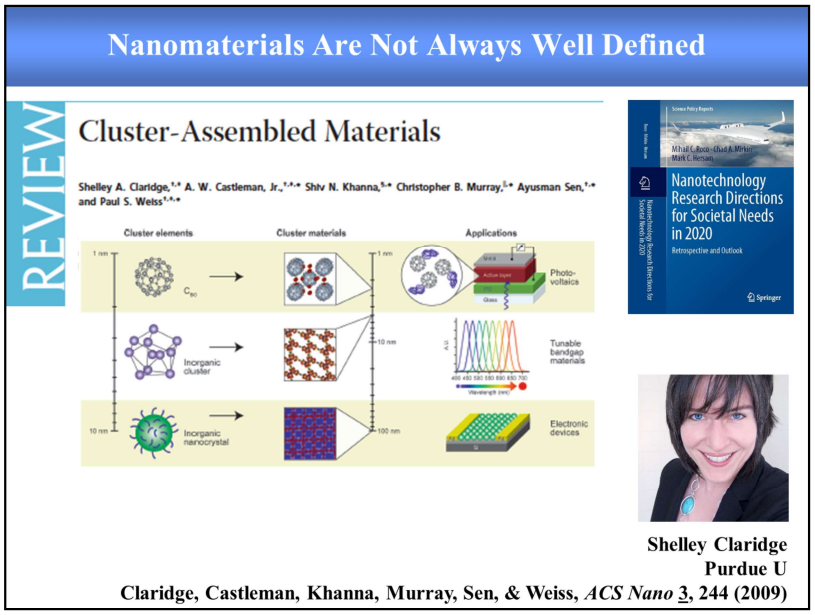
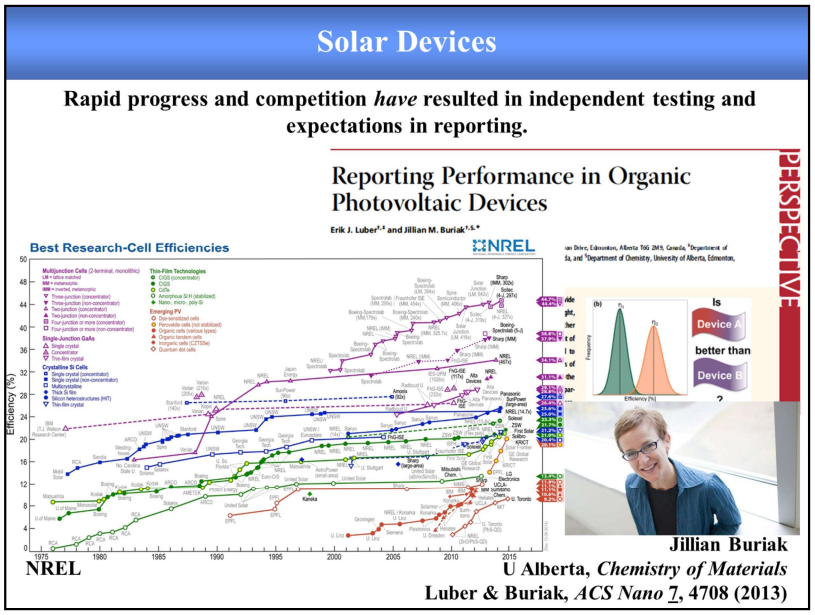
>> Lisa Friedersdorf: Good afternoon. This is Lisa Friedersdorf, Deputy Director of the National Nanotechnology Coordination Office. I would like to welcome you today to the webinar entitled "All Hands on Deck for Improving Data Quality." This webinar is the capstone event for our 2015 Nanotechnology Knowledge Infrastructure and Sensor Signature Initiative webinar series. Today I would like to welcome Dr. Paul Weiss, distinguished professor at UCLA and editor-in-chief of *ACS Nano* , and Dr. Ewan Birney, Director of the European Bioinformatics Institute. Their bios are available on the webinar site on nano.gov. I would like to thank them for participating today, especially since they're coming to us from Ghana and the UK, and we appreciate them making the effort across time zones to play a role in the webinar series. Paul will give an overview of data quality issues in nanotechnology, and Ewan will present a case study on how one community is proactively thinking about these challenges. And with that, I will turn it over to Paul. Thank you.



>> **Paul Weiss:** What I thought I would do is cover a range of topics related to understanding the precision of materials we work with and also being able to compare measurements--particularly devices--across laboratories. Defining these materials can be challenging since there can be disparity in the building blocks; it is not as simple as chemistry where we can define the particular molecules. Sometimes this can turn out to be an advantage.



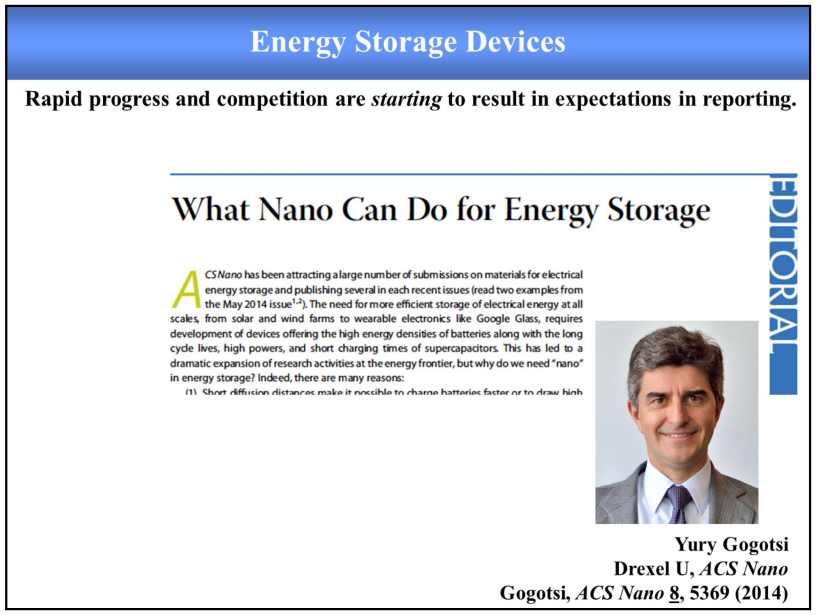
>> Paul Weiss: I'd like to start with this particular example because it captures many of these aspects. Engineering building blocks with materials other than molecules or atoms presents a unique set of challenges related to design and characterization. At the top of the graphic, we learned that we can make solids with clusters, such as fullerenes, as building blocks. This is more complex than the atoms and simple molecules that we usually think of as building blocks. At the bottom of the chart, colloids are used as building blocks to create a colloidal crystal. The shapes, surface interactions, and relative sizes of the components can already be used to control lattice symmetries and spacings with remarkable diversity, enabling us to pull out a set of interesting properties. In the middle example, inorganic clusters are used as building blocks, and you might imagine being able to tune the electronic, optical, chemical, and other interesting properties of materials. This is consistent with broader research efforts to develop the next generation of advanced materials, and relates in some aspects to the second decade of the National Nanotechnology Initiative.



>> **Paul Weiss:** In solar devices, there is a certification process for looking at the efficiency of devices that are being developed. But this doesn't mean that all the data that went into the development of these devices is being reported in publications. Dr. Jillian Buriak from University of Alberta and now the Editor-in-Chief of *Chemistry of Materials* published a very nice piece on what to expect in published work on organic photovoltaic devices. In this publication, Dr. Buriak discusses the statistical significance of the power conversion efficiency value and whether the observed change or improvement in performance is truly greater than experimental error. In my personal view, many publications come out as short communications without sufficient detail for anyone to follow up. We felt that this was a huge problem for the field, and that is why this high-profile, comprehensive article, which looks out over the field, was important to address those issues. In addition, the National Renewable Energy Laboratory (NREL) maintains a plot of compiled values of highest confirmed conversion efficiencies for research cells, from 1976 to the present, for a range of photovoltaic technologies (see chart). Overall, rapid progress and competition have resulted in independent testing and expectations in reporting as indicated.

Energy Storage Devices

Rapid progress and competition are *starting* to result in expectations in reporting.



EDITORIAL

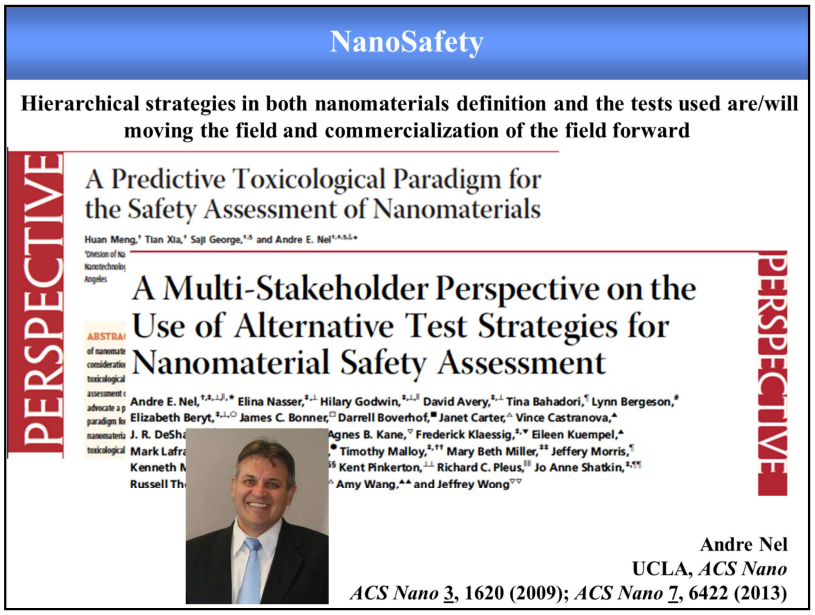
What Nano Can Do for Energy Storage

ACS Nano has been attracting a large number of submissions on materials for electrical energy storage and publishing several in each recent issues (read two examples from the May 2014 issue[1,2]). The need for more efficient storage of electrical energy at all scales, from solar and wind farms to wearable electronics like Google Glass, requires development of devices offering the high energy densities of batteries along with the long cycle lives, high powers, and short charging times of supercapacitors. This has led to a dramatic expansion of research activities at the energy frontier, but why do we need "nano" in energy storage? Indeed, there are many reasons:

(1) Short diffusion distances make it possible to charge batteries faster or to draw high

Yury Gogotsi
Drexel U, *ACS Nano*
Gogotsi, *ACS Nano* 8, 5369 (2014)

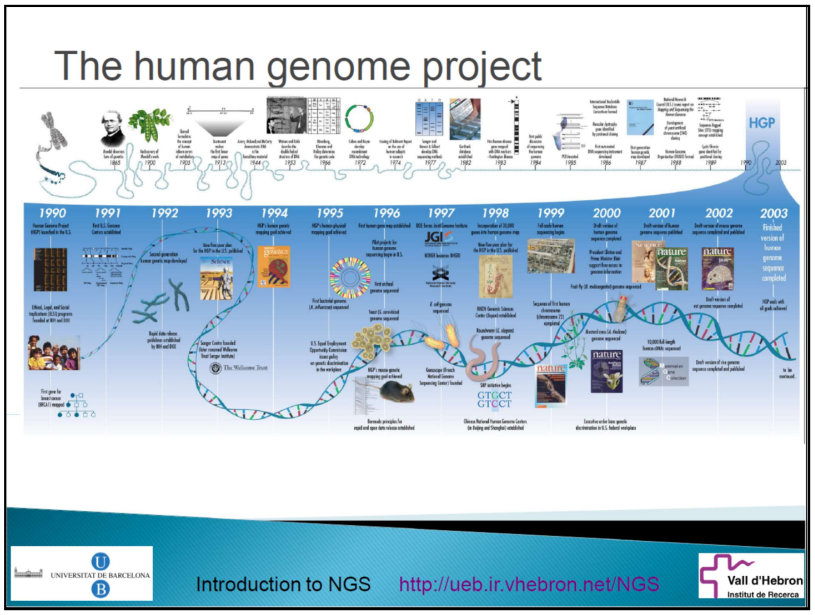
>> Paul Weiss: In this editorial piece recently published in *ACS Nano*, we tackled essentially the same idea with energy storage devices. Now there is more variation in materials, designs, and methods, and the field is growing rapidly. We anticipate, as the field progresses, getting greater and greater details and understanding of these devices. Currently, nobody does certification like NREL does for photovoltaic devices. There isn't a chart for energy storage like there is for solar cells (see previous slide). Instead, the editors at ACS Nano follow work in the field and request that authors identify and explain the mechanisms behind the reported performance of a given device and ensure that the electrochemical testing is done correctly.



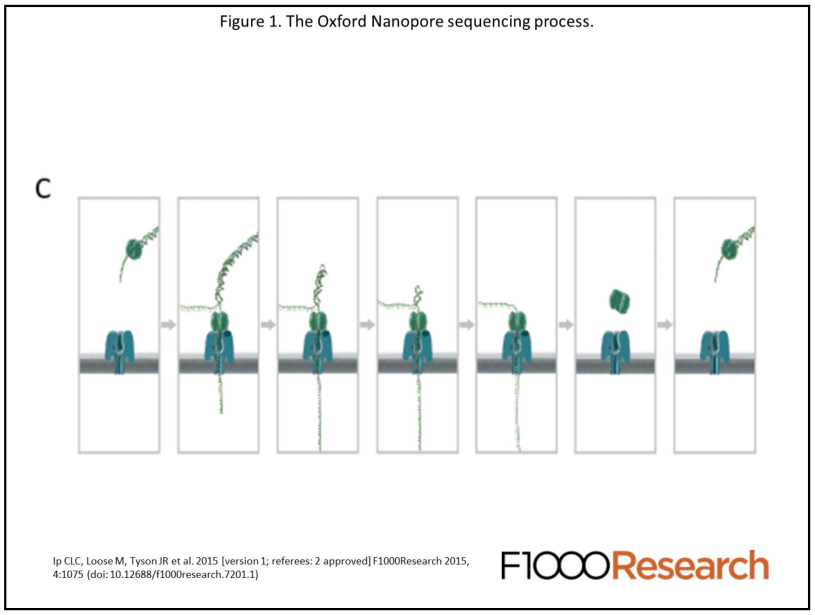
>> Paul Weiss: In terms of nanosafety, the law is such that in the U.S., nanomaterials should be treated as chemicals. One could argue that there are 100,000 or more nanomaterials out there; so it would be really difficult to conduct testing of all nanomaterials for comprehensive safety assessment. The rate of development of nanomaterials thus calls for the consideration of appropriate toxicological paradigms for safety assessment of nanomaterials. Dr. Andre Nel advocates for a predictive toxicological approach by establishing mechanisms and pathways of injury at a cellular and molecular level to prioritize screening for adverse biological effects and health outcomes *in vivo*. Furthermore, the use alternative test strategies (ATS) could help reduce reliance on animal testing through the use of *in vitro* and *in silico* methods. Such hierarchical strategies in both nanomaterials definition and the tests used could help move the fundamental research and commercialization of the field forward. In some cases, this approach could also help define a certain class of materials that could be moved forward for commercial applications. For instance, certain nanomaterials can exhibit positive interactions with biological systems, which could be tailored for therapeutic applications. This is a very rich area. The idea is to select certain classes of materials and go deeper and deeper as needed to further define their properties and potential applications.

>> Lisa Friedersdorf: Great. Thank you very much, Paul. We're now going to move to our second presenter for today, Ewan Birney.

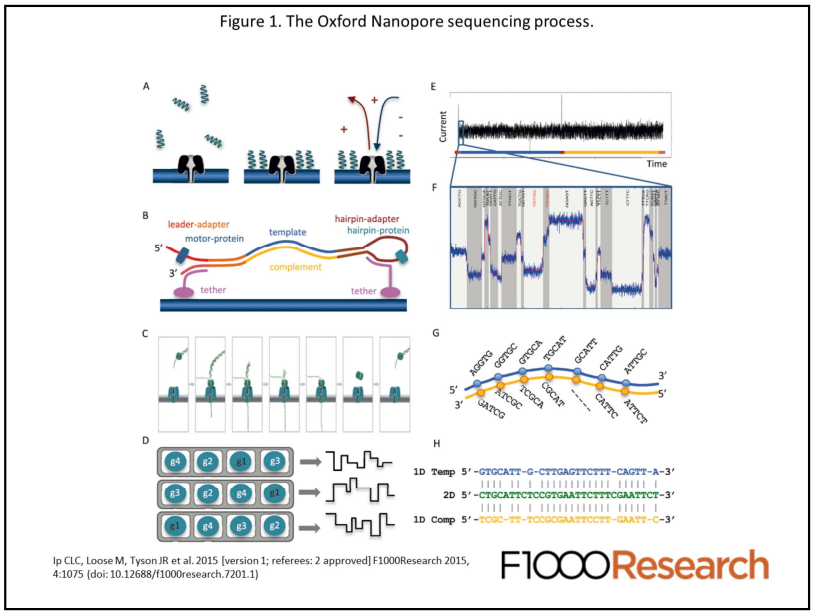
>> Ewan Birney: So just to start, I'm from a very different community here. I don't know how many viewers are biologists, but I'll be talking about a biological application of nanotechnology called nanopores. My first slides are just to say I'm Ewan Birney. I'm from EMBL-EBI, a European organization, and I am just south of Cambridge near the Sanger Institute. EMBL-EBI is next door to the Sanger Institute. I do have a conflict of interest; I'm a paid consultant to Oxford Nanopore. I have been providing advice to them for the past seven years, and I know them well. I'm also a long-time genomics and bioinformatics expert, and I have consulted for a number of companies.



>> Ewan Birney: The next slide gives you a sense of the Human Genome Project. The top line gives you the story from Mendel, which was a long time ago. In the '90s, there was the start of the automation of DNA sequencing. DNA is the fundamental code that is in every one of our cells. Every skin cell, white blood cell, and neuron has three billion bases of DNA letters, and DNA sequencing is about reading that code--not just for humans, but also for all of life. There's been a succession of instruments that have done that readout, and they started off being very manual. They became automated in the early '90s and went through a big change in 2005 with a couple of technologies. A dominant technology was Illumina, and what I'm going to discuss now is nanopore technology.



>> Ewan Birney: So in my next slide, I go and have a look at the basic concept of nanopores. So you have a lipid bilayer, and in the production machine (this is slightly different) there's a small hole piercing a membrane. These are biological holes; these are proteins that have these holes. Then the DNA sequence is attracted to it because of a voltage difference across the membrane. That voltage difference then also allows ions--water ions mainly--to go through the hole, and those water ions form a current. It is the change in current that provides the sensing of the DNA, and then a second enzyme called the motor enzyme steps through the DNA, releasing it slowly into the pore. One of the key things was getting this speed down. Nanopores were known to take DNA through the nanopore for a very long time, but it happened so quickly you couldn't sense at the same time. So a key component was slowing down the DNA as it went through the nanopore to create the ionic changes.



>> Ewan Birney: On the next slide there is a complete overview of this process. You have the nanopore sensing, and there is molecular biology to produce the right piece of DNA (section B here) ready to be entered into the pore. In the device, there will be any number, usually 500 or 400 active pores, simultaneously reading. Because everything is done asynchronously, there has to be a complicated piece of electronics. On the top, right-hand side you see a signal that comes from that nanopore. It looks jumpy, but zooming in, you can see very characteristic flat regions where there's a constant current, and it's pretty clear that those are individual points of the DNA sequence, which give a specific current. It would be great if they were just reading one of the four types of DNA bases (A, T, G, or C), but in fact they're not. They're reading five or six at a time, and so you have to deconvolute this signal. You use pretty standard signal processing for that, which is a hidden mark-up model-style framework. There are some details here that I am skipping over, but I will move on.

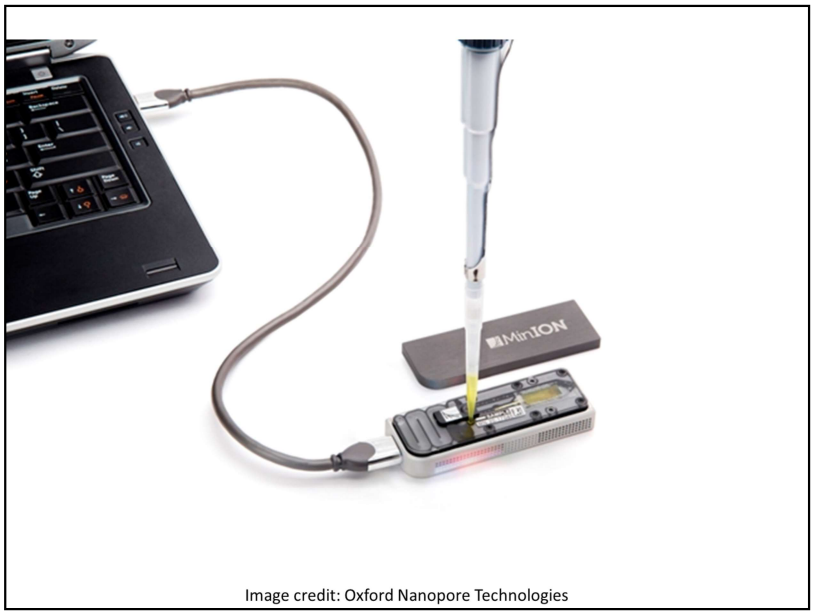
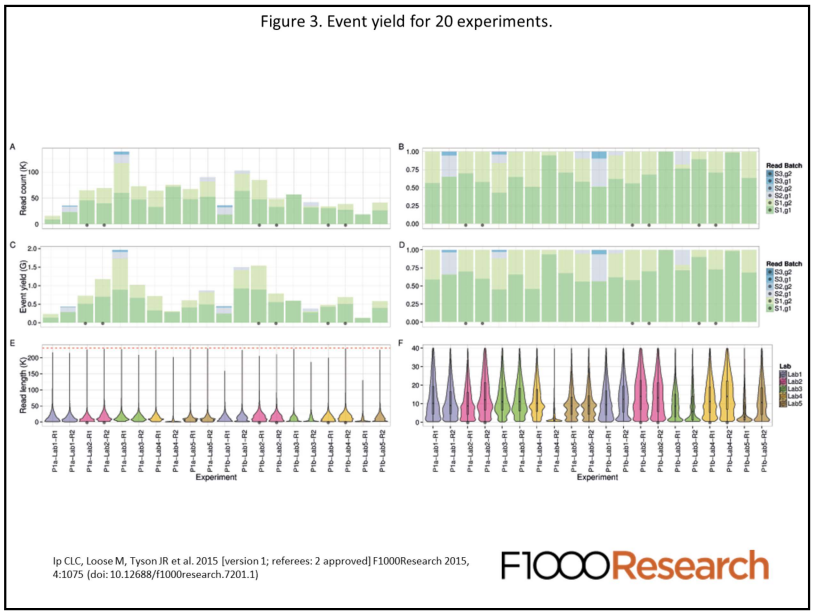


Image credit: Oxford Nanopore Technologies

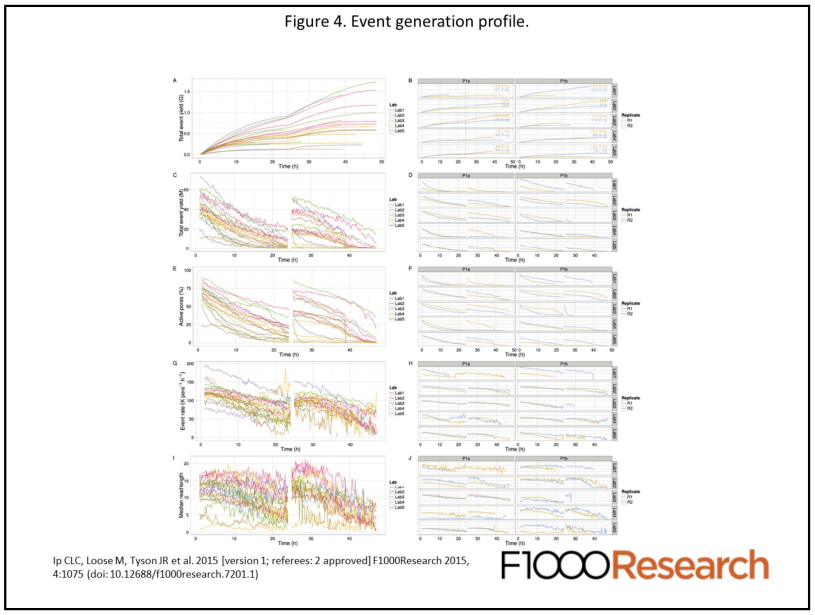
>> Ewan Birney: Oxford Nanopore has been working on this technology for a while. Slide six shows the device you can buy now, which is called a MinION, and the important thing here is portability. I'll come to that again in a moment. Just to mention, in 2016 a bigger device with more nanopores is coming along. Obviously, the actual sensor is a tiny little protein--a single molecule--and so really this is about how you want to shrink or not shrink the electronics and how you want to scale the electronics.

>> Ewan Birney: So why do people get excited about this? It is truly portable. The top left of this slide shows someone in Costa Rica sequencing a frog at the top of a mountain. On the bottom right is a far more practical application. These are people in Liberia sequencing Ebola, and you can see the sequencing device there just in front of that researcher on the computer. That has really transformed the diagnosis of Ebola in the field, moving it away from a ten-day process to a one-day process. Then NASA is exploring putting the MinION into space and therefore being able to sequence in the International Space Station. This would involve very practical things such as monitoring humans, but actually there are much more mundane things, for example, monitoring bacterial content in the space station, described as the microbiome. A final thing: if you're a sequencing geek, you will get excited for a different reason. Nanopore read lengths are seemingly only limited by library preparation. That's not totally true, but it's extremely clear that nanopores can read very, very long reads. And that's very exciting. What is the downside? The downside is that the error rate is not the same as some of the other machines, and that's somewhat in the software and somewhat just inherent in the process. But, the level of reads mean this is already a valuable DNA sensing device, and it's real-time as well.

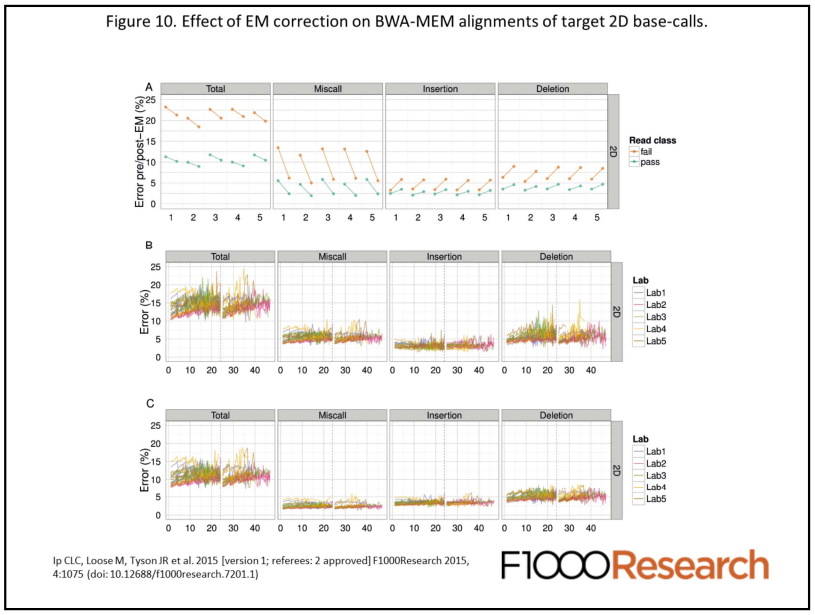


>> Ewan Birney: So now I want to move on to the MARC consortium. This is a consortium that I helped put together of scientists around the world to really test out the reliability and error rates of this machine. So the goal of the MARC consortium was to provide a consistent set of data with tightly defined protocols. When you're trying to understand what is going on in some of these devices, especially in a portable device, there is a lot of diversity of labs, samples, and preparation of all sorts. So just understanding reliability when you vary one axis is incredibly important. Then having provided some tightly defined protocols, we also wanted to, of course, improve things like read length and other things. A key thing is just getting out into the community the way to think about the data. So in all of these things, just the data shape is usually unique to the device. You're trying to measure the same thing as other machines, but the details are different. Just having a community understanding of that is very important.

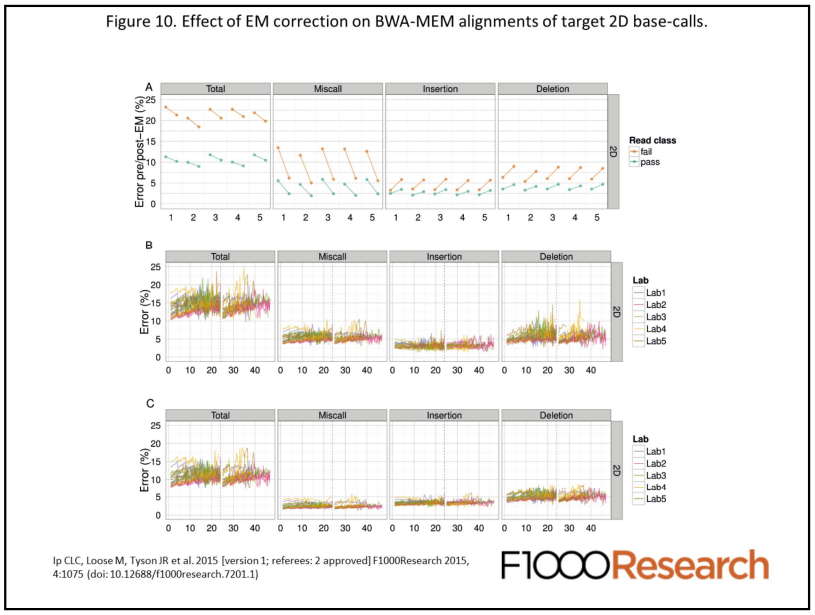
So I'm going to really be presenting things from this paper which we published in F1000. I'll mention that it is a rapid publication. The first author is Camilla Ip, and the team put together a really excellent paper. I'm an author on this paper as well, and most of my slides will come from this publication. Slide 13 gives you a sense of what we did in this consortium. There were five different labs that ran a total of four different experiments, so we have 20 experiments. This is one of the most basic readouts, which is the amount of data acquired in each experiment. You can see that it is quite variable, but the key thing here is that all experiments were at least somewhat successful.



>> Ewan Birney: On this slide, I don't really expect you to digest this figure, but clearly the top figure is showing over time the most important feature, which is yield. The next set of figures down breaks down why that drop in yield happens. There was something going on in these machines, and on the right-hand side we break it down by lab. Although it's quite hard to get statistics on some of these curves, I think visually one can really see that between-lab variance is higher than the intra-lab variance. That says there's a lab effect. So there are some details about how labs are preparing DNA, which is triggering this differential yield. The overall figures, by the time you get to the bottom, is probably the most inherent thing to the machine, which is that the length of the DNA going through the machine comes from sampling the DNA in the solution, and that looks pretty flat, actually, for most experiments. There's an interesting tile to the right, finally, but that is sort of reassuring.

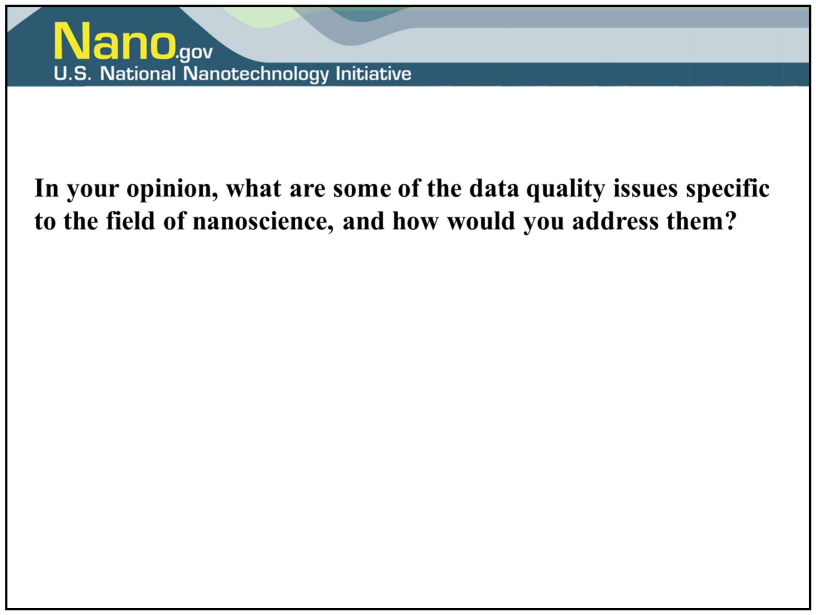


>> **Ewan Birney:** I just wanted to highlight another figure in this paper, Figure 10. This is showing some of the analysis we did. This is plotting the five different labs with the different points, each experiment, and then the connection between the points is unchanging--one of the analysis routines. What you can see in the left-hand panel is that there's always a slight downward trend overall. In the next three panels, which say “missed call,” “insertion,” “deletion,” you can see that really what is happening is that the algorithm is putting in less “missed calls” for a slight increase in insertions and deletions. Now, again, to explain this figure you’d have to know quite a bit about how DNA sequencing works and the output, but this is showing that there was innovation in the analysis precisely to understand the error behavior. This both improved the overall error rate, which was great because the error rate goes down, but perhaps more importantly, it gives us insight into how and why the machine generates errors.



>> Ewan Birney: So I just wanted to give you a sense of what this kind of approach is like. This is very much a community approach. We had a very broad call across the people using these machines. I didn't meet many of the authors, and I still haven't met them in person. I've only met them through teleconferences and email. But be aware, when you do this community approach, you always get many more lurkers than active people. So you usually have a 10:1 ratio of people who are active versus people who just come along for a ride, tune in, and look about. I think an open door approach is good, but discount your numbers by about a factor of about 10. The second thing, which is quite important, is that everybody wants to rush on and do interesting experiments. We started with very boring experiments, and I think that's quite important as well. In a community approach, you have to remind people to do these baseline experiments. Boring but necessary. Of course, you want people to understand and use the data. In molecular biology we have a great tradition of sharing data, and we share data immediately with the entire community on this publication. We've published with *F1000Research*, which provides extremely rapid publication. It's a vehicle before peer review, and the peer review happens very quickly. It's citable throughout. So it's citable before it's peer reviewed, and citable after it's peer reviewed. I can see many more pieces of analysis coming through this scheme.

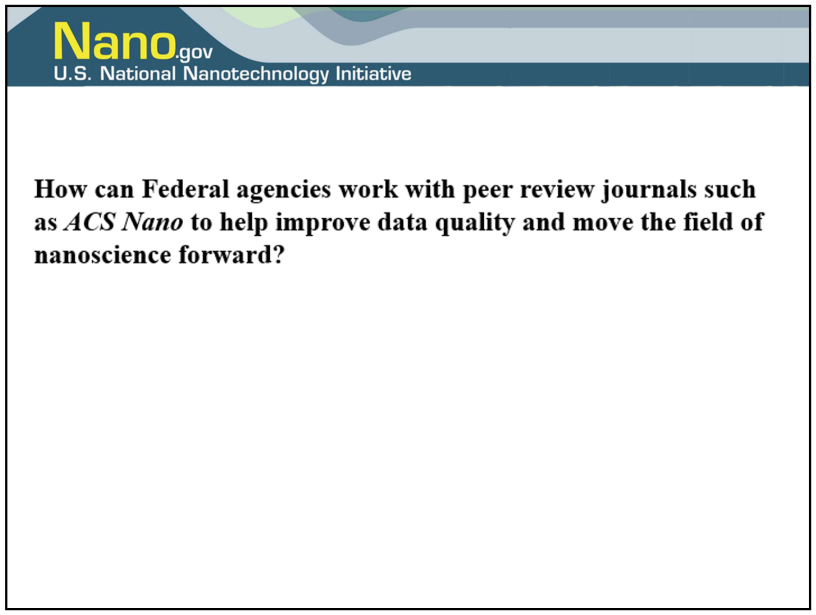
So my final thanks is to thank Oxford Nanopore both for making this technology, which I think has been transformative really, and also for being so open themselves in giving access to technical details as the consortium developed things. The second list of thanks are the five labs that did the experiments: David Buck's at Oxford, Mark Akeson's at UCSC, Sarah Goodwin's at Cold Spring Harbor, Hans Jensen's at ZF-screens in the Netherlands, and Justin O'Grady's at the University of East Anglia. Camilla Ip at Oxford and Miten Jain at UCSC were the two lead analysts for this paper and really pulled it together. I hope that was useful. I'm happy to answer questions.



>> Lisa Friedersdorf: Thank you so much. That was really fantastic. I appreciate your comments this afternoon. They're very applicable to the conversations that we've been having with the NKI community, and I think this was very, very helpful to our communities. So I'm going to move to the questions now. We have received a number of questions, and I would like to encourage participants to continue to submit their questions. The first one is: *in your opinion, what are some of the data quality issues specific to the field of nanoscience, and how would you address them?*

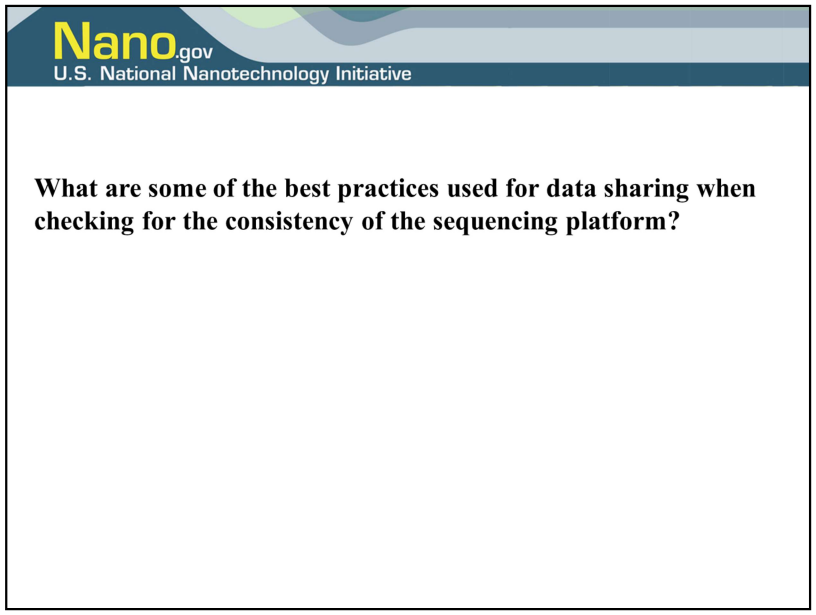
>> Paul Weiss: Very good. As I said earlier, it's really a *range* in nanoscience. Some materials we can define very precisely. Others, much, much more loosely, if we haven't yet developed the tools we need, or in some cases we even have dispersions, either intentionally or because we can't yet make more precise materials. We really have to deal with those variations, and that goes differently in the different areas. Materials characterization is one issue. One of the other big ones, as I said, is making sure that we're reporting in such a way that people can follow up on what has been done in terms of synthesis, assembly, and fabrication, as well as characterization.

>> Ewan Birney: I think it's a very generic question in the sense that you have to be driven by the data type. In some sense, I don't think there's something fundamentally different between nanoscience and other sciences in trying to be reliable and thinking about data quality. But a lot of it is fundamentally about measurement. In measurement, things like nanoscience have a difference in the sense that you are going to be dealing with single molecules, you may well be dealing with quantum effects, and you have to be able to model that in your measurements. So the question may evolve from "is there a difference in data quality issues?" to "what are the specific ways you could measure things?" If those measurements have inherent error, for example, due to molecular fluctuations or quantum fluctuations, one needs to take them into account when thinking about data quality.



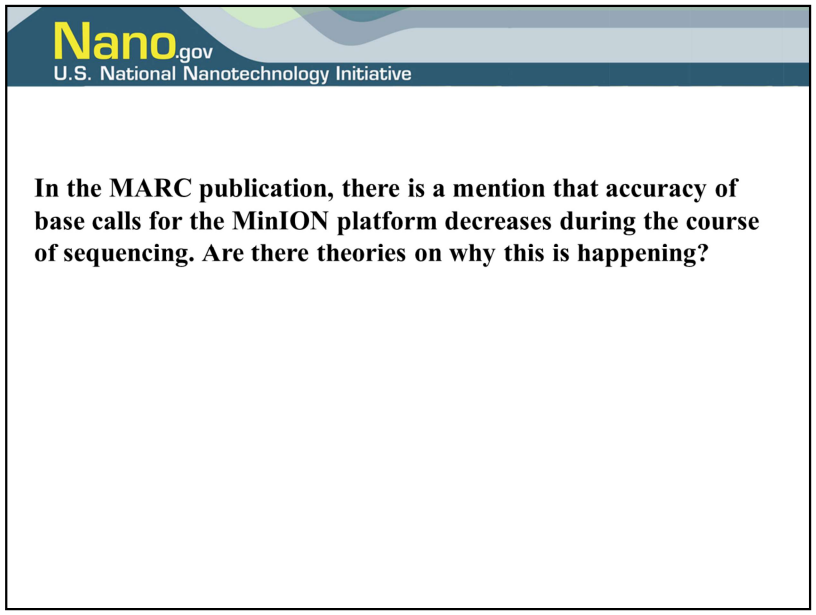
>> Lisa Friedersdorf: Great, thank you. The next question is: *how can Federal agencies work with peer review journals such as* ACS Nano *to help improve data quality and move the field of nanoscience forward?*

>> Paul Weiss: I think we have an example of that kind of outreach right here with *ACS Nano*. You know, we've made an effort to guide the field in this way, and I think that's why you came to us. We've also tried to more generally address these issues globally —not just in the U.S.—by going through each area where we see opportunity to do better characterization, and lay out the current state of the field—and advance that over time. We don't expect this issue to be static. We're doing our best to continue in our efforts to guide the field of nanoscience forward, and we see that as our role.



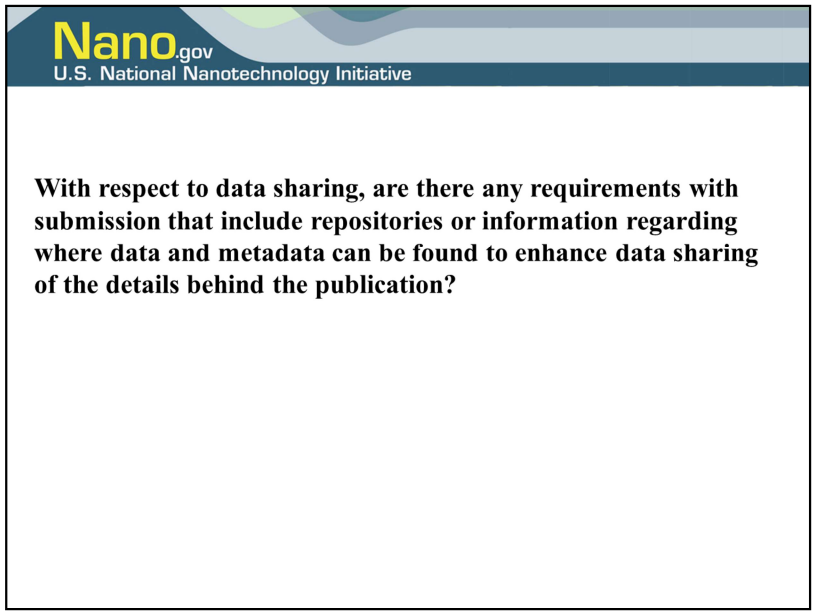
>> Lisa Friedersdorf: The next question is related to the MinION Analysis and Reference Consortium: *what are some of the best practices used for data sharing when checking for the consistency of the sequencing platform?*

>> Ewan Birney: I think one of the best practices is to be open about your data and let many people participate. There is always a question about how you structure that. For ourselves, we are very open inside the consortium, and we, as a consortium, decided to publish data quickly. We went from data generation to publication within five months, I think, which is quite remarkable. The other thing, of course, is that you have to standardize your metadata and make sure you have a central place where both data and metadata can be accessed. That's obviously just straightforward good practice, but I'm always amazed at the number of places that don't do that.



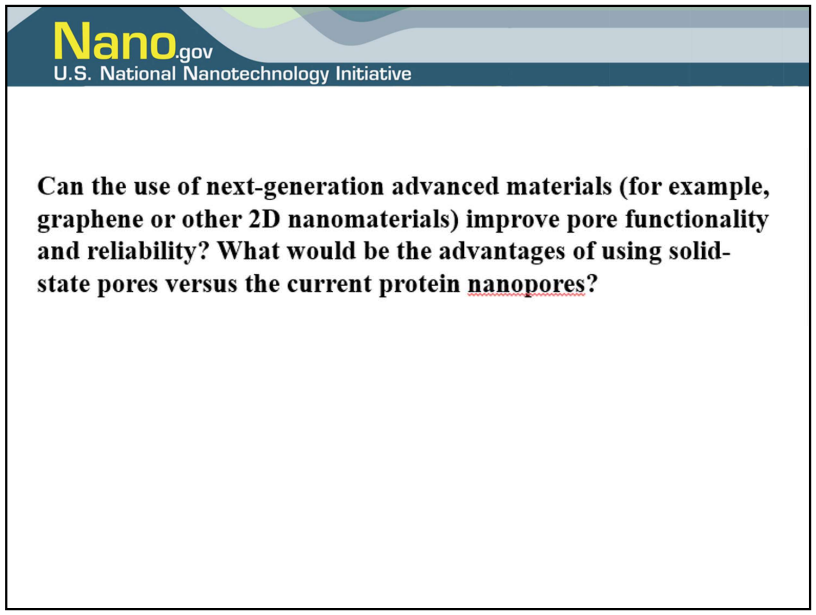
>> Lisa Friedersdorf: That's certainly an issue that I hear discussed a great deal, and it's a challenge that I think we as a community need to continue to work on. The next question is: *in the MARC publication, there is a mention that accuracy of base calls for the MinION platform decreases during the course of sequencing. Are there theories on why this is happening?*

>> Ewan Birney: Yes, there are many theories. One of the slightly frustrating things about the MinION platform is that the team at Oxford Nanopore is innovating about at the same rate that we are analyzing. Slightly frustratingly, they have moved the goal posts forward, and we know this has been on their radar to understand this. One theory is that there may be some sort of rare pore blocking event that may cause the error rate to tail off. By having quite an open interface with the company, you get insight into their pipeline, and I think the company feels confident they may have solved that. Now, the key thing is that the community needs to digest new techniques. What I would stress there, beyond the theories, is that we have a way of thinking about the data and a way of plotting and analyzing the data. Therefore the key thing is whether the error rate or error profile is good enough to do your science. That's true now. But there are other bits of science we want to do, and that will need better control of the error profile. Going to your point, I think the thing we're doing there with the MARC consortium is to make the analysis techniques, publish them, and make them open, because that community knowledge of what a good run looks like for a nanopore is key.



>> Lisa Friedersdorf: Great. Thank you. The next question is: *with respect to data sharing, are there any requirements with submission that include repositories or information regarding where data and metadata can be found to enhance data sharing of the details behind the publication?*

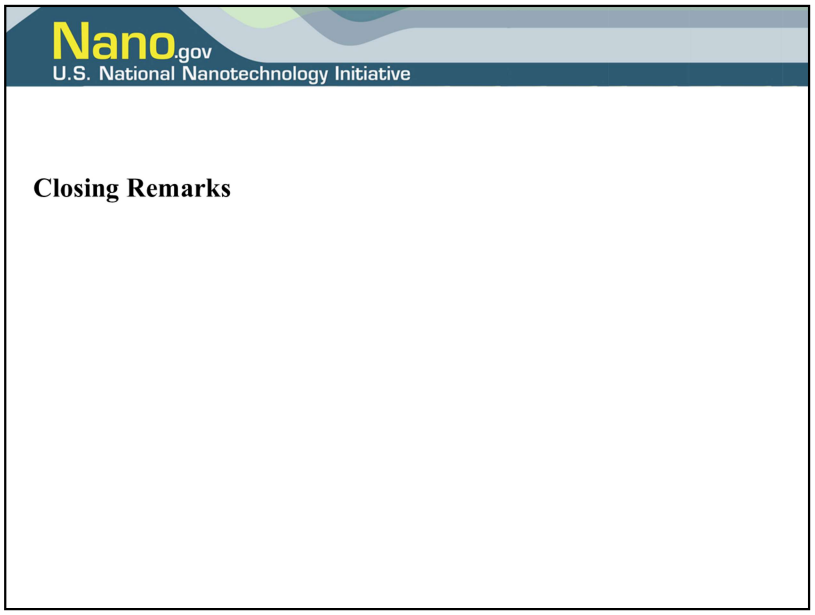
>> Paul Weiss: There are significant requirements in terms of what needs to be included in paper and supplementary information and making materials available and so forth. *ACS Nano* has made an effort through the requirements that the authors are going to put in everything necessary and then work with other people who want to followup when possible to do more on their work.



>> Lisa Friedersdorf: That's great. Thank you, Paul. Our final question in the queue is: *can the use of next-generation advanced materials (for example, graphene or other 2D nanomaterials) improve pore functionality and reliability, and what would be the advantages of using solid-state pores versus the current protein nanopores?*

>> Ewan Birney: There's a lot of excitement and interest about graphene and solid-state pores. Clearly, there's the opportunity there for both reliability and robustness. I think most people are interested in the robustness aspect. In other words, most protein pores are quite consistent, which is a question on the manufacturing side. There's no doubt that something which is more like a thin material, ultimately graphene perhaps, should provide a more robust pore solution than a protein inserted into a bilayer. To be honest, I think it's at least a couple years away, if not more, having seen what one has to do to get a protein pore working. Don't forget you've got this problem as well of controlling speed, and that may be quite hard to handle in the context of a solid-state pore. That's why it's exciting to do research!

>> Paul Weiss: I'll jump in to say, there are ways to make pores in graphene and probably other 2D materials precisely, including enabling precise chemistries inside the pores. That's an area of very active research in a number of labs right now.

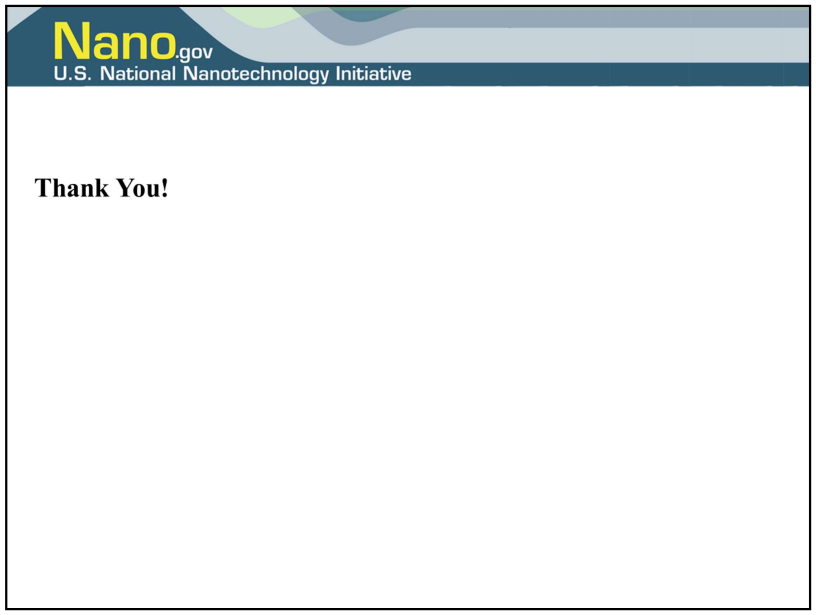


>> Lisa Friedersdorf: Great. Thank you both for answering that question. That takes us to the end of our prepared questions, but I would like to give you each an opportunity to share any closing thoughts that you might have on this topic.

>> Paul Weiss: I think this idea of getting key components in different areas together is critical for us to move nanoscience and nanotechnology—not just the research—forward, but get to the point where we see things like nanopore sequencing and much more taking advantage of the tremendous investments that we've made around the world in the field. I urge the listeners to go ahead and write to me or other editors in the field if they see an opportunity to advance a particular area by saying, "okay, here is something that is missing in publications, talks, or other areas, or here is a good figure of merit we can use to do comparisons between one laboratory result and the next." By all means let us know. and we'd be interested in following up.

>> Ewan Birney: I think there's a very, very rich opportunity in the interdisciplinary space between molecular biology and nanotechnology. I think nanopores are just one example in a large space. Molecular biology really is evolution's nanotechnology. Every time I've interacted across a disciplinary boundary, new ideas have popped up that have been absolutely transformative, often for both fields if not just on one side. So I would like to encourage the listeners to talk to molecular biologists and spend a little bit of time learning their language, just as I encourage the molecular biologists to hunker down with some chemists and nanotechnologists.

>> Paul Weiss: I couldn't agree with that more. Actually, let me point people to look for a big piece in *ACS Nano* on a technology roadmap for the microbiome, which is coming out in a few days. It's not quite out yet, but it's just about to get into production. It is one example, as Ewan said, that is taking the opportunity that is provided by the functions of biology being at the nanoscale. That's a tremendous opportunity that we don't want to miss.



>> Lisa Friedersdorf: Great, Paul. We'll keep an eye out for that. Personally, I would like to say that I think the interdisciplinarity is what really brings nanotechnology to where it can be so powerful, and I think that's what makes it so much fun. I would like to take the opportunity to thank both of our speakers and our listeners today. I feel like the insight that the speakers provided is very valuable to our community. I know it certainly will play a role in both our sensors and our informatics Nanotechnology Signature Initiatives. We will provide the transcript and the slides on nano.gov as soon as possible. Thank you.