Emerging Genetic Technologies and Their Medical and Public Health Applications

Nicholas C. Dracopoli, Ph.D.
Vice President of Clinical Discovery Technology, Pharmaceutical Research Institute, Bristol-Myers Squibb

---

DR. McCABE: Our next talk is emerging genetic technologies and their medical and public health applications by Dr. Nicholas Dracopoli, who is Vice President of Clinical Discovery Technology, Pharmaceutical Research Institute, Bristol-Myers Squibb.

DR. DRACOPOLI: Thank you.

Firstly, I'd like to thank the Committee for the invitation to present. It's an honor to be here today and to see many old colleagues on the Committee and in some of the surrounding groups.

What I'd like to do today is really to focus my talk on the perspective, at least from a pharmaceutical company, on the development and impact of pharmacogenomic methods and applications and the profound impact that those are starting to have and I think will progressively have over a period of time.

The questions I'm really going to focus on, and this is really a technology-driven approach, is what genetic technologies are appearing on the horizon that we're using in the discovery areas right now, and how these are going to be transitioned from a discovery environment into a clinical lab testing environment, and to just discuss some of the issues that are going to be raised by the development and integration of these technologies for the delivery of therapeutics. Then also, finally, to just have some suggestions on the need for public policies that need to be in place to facilitate the development of pharmacogenomic approaches for drug development.

Firstly, a definition. What is pharmacogenomics? We view it as a very broad definition, but it's the use of markers of biological variation, and that can be at really any level -- DNA, RNA, or a protein -- to predict a patient response to pharmaceuticals, and that can both be to predict drug efficacy as well as also predicting adverse events. Secondly, also molecular pathology in terms of more effective definition of disease so that we can identify disease subtypes that can then be treated more effectively and uniformly with the knowledge of that underlying disease heterogeneity.

In a sort of cartoon for pharmacogenomics, the idea is that you can basically pre-identify individuals before the selection of therapy. So you can reduce the risk of adverse events, as well as increasing the proportion of patients who would benefit from a therapy. So the profiling method -- here we describe it as a microarray, but it could be any profiling method, whether it's proteomics or even high-field nuclear magnetic resonance for looking at metabolites, or SNP gene testing -- is to take a heterogeneous group of patients who in this cartoon have a 60 percent benefit of therapy, pre-screen them into two groups, and then you can enrich one group who you would treat with a higher efficacy, and similarly identify a group of patients who are least likely to respond who you could then select for alternative therapy.

Now, the need for this really varies according to the type of indication that you're trying to treat. For example, if there isn't a simple means of measuring the response of your drug before you treat a patient, then this test could be really useful. So in areas of oncology, for example, understanding the particular subtype of a tumor and identifying the individuals most likely to respond is something we have no way of

doing right now. We treat a patient and we see their response, and then we put them onto a second round of different therapy or a third line of therapy, depending on how they progress with the disease.

But if you're looking at, say, for example, drugs for reducing lipid, drugs for reducing blood pressure, we have a fairly immediate output for whether the drug is working or not working. So you could argue that it's much less important to develop tests in those areas where there is an immediate output.

So why do we need pharmacogenomics? Well, the bottom line is that in many of our therapeutic areas, the efficacy of our drugs is very low. My background is primarily in the oncology area where this is particularly the case. We're seeing now, for example, at the recent ASCO meeting huge excitement about new drugs for colorectal cancer, which realistically have a 10, 15, maybe 20 percent response rate and are extending life only a matter of months, and these incremental changes, while huge, are sufficient to have headlines in major papers across the world.

So I think there really is a huge area of opportunity for increasing the efficacy of our compounds, and I think the underlying scientific reality is that these diseases are highly heterogeneous and it's very unlikely that individual therapeutics are going to broadly impact all types of a particular cancer. So we need to be able to identify the molecular pathology underlying these diseases in order to increase the efficacy of our compounds.

To look at cancer for a moment in particular, we're taking a sort of technological perspective. We need to look at two types of underlying causes of variation in response to a drug. These are both driven through the host or these are the inherited factors that essentially are determining the individual's response to the drug, or pharmacokinetic differences, and then also looking at the tumor, looking at the genetic changes that are occurring within the unstable tumor to look for predictive markers and patterns through profiling approaches that then can predict individual response.

These two different methods require totally different technological approaches. So, for example, looking at germline variants, as was mentioned earlier, looking at things like the P450 genes, they are known genetic variants and can be tested for by a relatively simple genotyping test for single nucleotide polymorphisms, and we're developing public databases that are basically helping us to understand how these genetic variants impact an individual's response to the drug and the pharmacokinetic differences, and we now have examples in the clinic where drugs are being dosed based upon genotyping or drug metabolism or drug transport genes. Mercaptopurine treatment for childhood leukemia is one good example.

I think one of the areas of really promising research that we've seen over the last couple of years has been in profiling approaches of tumors where we're now starting to correlate individual gene expression profiles in tumors to drug outcome. So the hope is that we can use these approaches, both with broad-based expression profiling methods as well as analysis of known mutations, to class groups of otherwise homogeneous cancer patients into multiple classes and then ask if those individual classes have different responses to different therapeutics. I think this is an area that we're now starting to see and will soon, I think, have a significant impact on the clinic.

To look at the range of technologies that we're looking at, they really fall into three major categories. This is, again, looking at protein or DNA or RNA. I think the important thing in looking at these technologies is from a pragmatic and cost-driven perspective. The RNA-based approach or transcriptional profiling is still the only approach that allows us essentially to look at all of the known messages, or nearly all of the known messages, in any particular sample.

If we look at proteomic-based screening of plasma, then we can look at complex mixes of up to 100,000 proteins. But a single mass spectrometry experiment can only resolve 2,000, maybe 3,000 complex protein mixes. So in order to be able to scan a proteome, we have to look at a large number of experiments essentially driven by liquid chromatography or other affinity-based methods to break up the proteome into fragments and then look at it in groups of 2,000, 3,000, 4,000 proteins at a time. So here, these experiments are often hypothesis driven, and we have to select particular types of protein, particular targets to look at in these experiments.

In the genotyping world, we now, through one of the great benefits -- one of the immediate benefits of the Genome Project has been the discovery of massive numbers of SNPs. There are now more than 2 million SNPs, I think, in the public domain.

DR. LANDER: Four million.

DR. DRACOPOLI: Four million. Sorry. I don't want to underestimate.

DR. LANDER: That was last week.

DR. DRACOPOLI: Last week, right. There will be 6 million next week.

So more than 4 million SNPs in the public domain. Here, the issue is not so much the discovery and knowledge of the SNPs; it's the ability to type them. The cost to type them is going down exponentially. But still, in order to type that many numbers of SNPs, we have to take approaches towards pooling DNA from cases and controls and comparing them. But the cost of doing massive genome-wide screening at the level of resolution that we need to do for effective association studies is still too high, and it's certainly too high to be applied to individual patients.

So I think the transcriptional profiling approach, in essence, is still the only approach we can look at that looks at basically all, or nearly all, of the known messages in a particular sample.

In order to be able to profile things effectively and actually have an interpretable result, we have to assume that the patterns that you're seeing within biological samples are not random and that there are a relatively limited number of patterns that you can find. So in this sense, this is a diagram that's sort of known as the circuit diagram for cancer, and I think the key thing why I often use this slide is that little yellow box in the middle of the nucleus there, which may be a little hard to read. It says "Changes in Gene Expression."

In essence, what this diagram is implying is that whatever changes are occurring in a tumor cell, whether it's changes in signal transduction, changes in regulation of apoptosis, changes in the regulation of the cell cycle, ultimately they will have at the end of that pathway an impact, and there are a relatively small number of pathways, and those changes, where they occur, whether it's in Ras signalling, whether it's in the cell cycle, control through mutations in the cell cycle or in other pathways there, there will be a limited number of events and pathways that are changed, and when they're changed, they will give us a consistent pattern that we can detect in cells that are taken from patients at the time of the biopsy.

So the profiling approach -- and this is true whether it's SNPs or DNA or a protein-based approach, even a metabolite screening-based approach -- is essentially to look for patterns. It's a massive pattern recognition problem. So in this cartoon you basically have grouped two sets of samples which are sensitive or resistant to a drug in this case, but you could use any other biological endpoint in which you're interested, and then in this case basically it's genes, but it could be proteins or SNPs. So in essence

here, you essentially have a random pattern of expression, and if you then imagine bringing in another tumor or another column, there's nothing predictive. There's no pattern that's suggestive of sensitivity or resistance. So, in essence, there's no marker.

So what we do in a profiling experiment is instead of having 12 lines here, you've got 30,000 lines if you're looking at an RNA profiling experiment, or thousands or tens of thousands of SNPs. You can imagine looking at that. You're essentially looking for a pattern where you then, through mining the information, look for sets of genes whose expression is most correlated to the particular biological endpoint you're looking for. So here in this example, you can see that there's a group of genes which are typically overexpressed in the sensitive cells and underexpressed in the resistant, and vice-versa above.

So you can imagine that if you now sample another tumor, line it up against here, you can predict does it line up with the left or does it line up with the right. If you do this often enough and you have enough samples, you can eventually get to a point where you have some statistical predictability based on these types of analyses. Clearly, there are numerous examples in the literature now where it is clearly showing that the fate of a tumor seems to be, in essence, hard-wired relatively early in the tumor evolution, and at the time of biopsy, at the time the patient presents with a malignancy we can actually take these cells, analyze them and get this sort of data that will help us really direct therapy much more effectively than we do right now.

The difficulty, in order to bring this dream to practice, is going from profiles to assays. The technologies we have available to us now are largely being driven by and developed for the discovery market. They're focused on delivery of very large numbers of markers on relatively small numbers of samples. Chips, bead-based arrays, proteomics screening are all basically focused on looking at many, many different analytes or markers on relatively small numbers of samples in terms of a discovery process. What we need to be able to do is to convert those profiles and the markers within those profiles that give us a prediction of a clinical output to assays that can be delivered in a clear, certified clinical laboratory environment.

There are two ways that this is really happening, in essence. There is going to be the evolution of these very large chip-based approaches, discovery tools, to focused arrays that can be delivered in the clinic, and also the development of specific clinical assays that are focused not necessarily on the very large numbers of analytes but to deliver these in the clinic. This is still the major technology issue, how we take these profiles, how we take these proteomic scans, some of which have been done here at the FDA and the NCI, how can we bring those to large numbers of patients in the clinic.

In a drug discovery perspective for finding markers, we basically go through a filtration process in preclinical and Phase I, looking essentially at a full array of all expressed genes to the point where we can prospectively identify sets of markers in our Phase II studies. The goal is that when we move forward into Phase III, we can have sets of markers that can be prospectively tested in Phase III studies to determine the correlation of the marker to the outcome. The hope is that in the end you will have a relatively small number of markers that can be converted into clinical assays. The numbers of markers that we need, the numbers of analytes that we need to test are going to be really critical in the choice of technology that's going to be used to deliver this.

Currently, if you look for example at Herceptin or Genentech's Susceptin, there you are measuring the expression of the HER2 protein in order to decide whether to basically add Herceptin to the standard chemotherapy regime you would use for those breast cancers. You're measuring a single analyte. In essence, you're measuring the status of the drug target. But now with these profiling methods, we are

developing 50 maybe -- some reports in the New England Journal of Medicine have said you need as many as 70 markers to have the full sensitivity. Others have said many smaller numbers.

The difficulty is how do we deliver these assays to standard immunohistochemistry-based approaches on tissue sections, and how do we basically bring these quantitative measures across multi analytes into practical application?

To give you one example from our own labs, these are in vitro prediction results, but this is drug response to Taxol. You can see two groups. The one on the left -- I think I'm getting this right -- is responsive, and the one on the right is resistant. You can see very, very different clearly predictive patterns that we can use, and then actually a quantitative measure of the expression. So if the genes on the left have relatively low expression, the genes on the right are highly expressed. It gives you a strong prediction of the likely response of those cells, and vice-versa for the opposite response. These examples now are becoming more and more common throughout the literature and being able to use this information.

Another way of using this information that doesn't involve bringing forward a test for efficacy or adverse events to the clinic is use of this data actually in the drug development process for decisions made during development. One of the key decisions is actually defining dose. Typically, an oncology drug is developed so that the dose that you move into a Phase II study is, in essence, the maximum tolerated dose. We usually don't know the biological effect of that dose or of lower doses at the time we take that drug into Phase II in the clinic.

The hope is by using these profiling approaches that we can identify markers that have a dose-dependent change relating to gene expression. They tell us, then, if we can analyze biopsy material, if the drug is hitting the target, and what level of drug you need to hit the target. If we can then basically correlate the impact on the target in surrogate tissues to efficacy results, we can then maybe bring forward drugs at a much lower dose than the maximum tolerated dose but a dose that is defined by biological efficacy rather than the adverse event or the toxicity profile. That, hopefully, given that many compounds fail because of these excessive toxicities, this may be a way that we can enrich the numbers of compounds we bring forward out of the early clinical experiments.

So the status of diagnostic testing right now is the paradigm of the Herceptin model, where you have a single-analyte test. The diagnostic test is actually the status of the drug target itself. It's used for a single indication and uses existing paraffin formalin-fixed immunohistochemistry technology. The technology is going to evolve to microarray or multi-analyte-based tests where the diagnostic test is not necessarily the status of the drug target but it could include it. So, for example, imagine the EGF receptor, anti-EGF receptor compounds. Even though there's no good correlation, the EGF receptor expression and drug response, understanding what's happening in that pathway may be key to predicting the response of those compounds. These tests could potentially be used for multiple indications, but they will require new diagnostic technologies and the routine preservation of RNA and nucleic acids, which is not standard clinical practice.

This will eventually lead to a reclassification of disease. It won't replace what we do now. Currently, at least in oncology, we classify disease by the tissue of origin. It's a lung cancer with a certain stage and grade, and then we treat based on the stage and grade of that disease. Here we will add on to that standard staging and grading molecular profiling information. It can be specific mutations of known oncogenes or known tumor suppressor genes. It can also be profiles, which will basically group patients into different classes. So we will build molecular pathology on top of the existing clinical and pathological information with these diseases.

So to specifically answer those four questions that were addressed at the beginning, what are the genetic technologies for health care and public health that are on the horizon? I think clearly we need to look at the biological paradigm of going from DNA to protein, and I think now adding metabolite analysis in there. I apologize to Eric for underestimating the number of SNPs, but that changes constantly.

The genotyping tools that are available --

DR. LANDER: You overestimated the number of genes.

(Laughter.)

DR. DRACOPOLI: Okay.

DR. LANDER: That only emphasizes that it's a moving target.

DR. DRACOPOLI: It's a moving target. But I'm within orders of magnitude, except for the SNPs, where I'm out by a factor of 2.

So the now 4 million known SNPs, the technology really has to evolve in the discovery labs to be able to screen subsets of those to the level that we are able to do effective association studies and then identify subsets of those that have clinical impact and clinical predictive power that can then be delivered in a different platform. The transcriptional profiling approach has now become very mature. It's widely used, it's highly standardized, highly reproducible, and we're seeing significant approaches to developing subset tools using this approach for specific clinical answers.

Again, for proteomics, we can look at complex protein mixtures, but we still cannot look at them in a universal way. They have to be subdivided.

Then finally, I thought I'd mention RNA interference as a new target that will eventually, I suspect, have clinical applications. But it was an enormously important tool for identifying and validating drug targets, but of longer-term impact.

What is the impact of these technologies on health care and accessibility and affordability? I think clearly that probably the most important thing is better definition of disease causation, which will lead to better therapeutics. We've already seen that in some cases. Gleevec was mentioned earlier. I think that's a wonderful example of basically understanding a type of leukemia that has a single type of mutation that's causative in the majority of those cases. Here there was a targeted therapeutic approach, and we're going to see more and more of that as we understand the mutations and pathologies underlying different diseases and the evolution of targeted therapeutics.

But targeted therapeutics implies that we have to know what the target is. So those are going to evolve with specific tests that we have to be able to bring forward in the market.

The molecular definition is easy. It's going to have impact on the pharmaceutical industry and more broadly in the sense of market segmentation. This is clearly an issue, but I suspect it's an issue that is rather overplayed.

The industry already sells drugs into a market that is already medically segmented. We sell antihypertensives, antilipids basically to all comers who have a high lipid profile, high blood pressure. They work 40, 50, 60 percent of the time. Those patients in whom those drugs don't work stop taking

them after a while. In essence, by definition, the market is already segmented. It wouldn't really hurt in some sense to be able to identify up front those individuals, but I don't think that's the key issue.

The key issue is to be able to identify up front the choice of therapy where it's really going to have an impact. Failure to treat a cancer patient with the optimal therapy is first-line. It results in a more systemic, more aggressive, more drug-resistant disease. There is an enormous impact if we don't get it right the first time. There's less of an impact if we give somebody a statin whose lipid levels don't go down. The drugs are essentially broadly safe, recent withdrawal of one drug not withstanding. But there isn't a huge adverse event. You measure the response and you can see if the drug works.

But clearly in oncology, this is an enormous unmet need and can have enormous and immediate impact on treatment of cancer.

We will identify new indications. This is the other side of the coin for market segmentation. We can basically, by understanding molecular pathology and common events across different types of tumors, we will be able to develop a drug, a drug will emerge into the market as a third-line indication for a relatively rare type of tumor and then will spread to more common tumors that have similar underlying molecular pathologies. Again, Gleevec is a good example. It was brought on the market for CML but clearly has application in a relatively rare gastrointestinal stromal tumor or gist tumor, which is a totally different type of tumor, has a mutation in a different but very related tyrosine kinase gene, and you can now see how that drug is moving from the original indication to a broader indication, and will probably be able to move to other indications where having mutations and tumors driven by those particular tyrosine kinases.

So there are swings and roundabouts. You have segmented markets, but also possibilities for new indications.

Finally, increased efficacy will be enormously important. As we go forward, it's very difficult to bring drugs forward that only have a 10 or 15 percent efficacy. It's enormously easier to run clinical trials and demonstrate efficacy if you have a higher rate of efficacy. We can do smaller studies, and it will be much easier to get approval. So we hope in some sense that a real balance in increased efficacy will result in lower attrition during our compounds developments.

Finally, better diagnostics that we will create, subsets of common diseases which will be shown to be non-responsive to existing therapies. That's an issue that we have right now. Those patients are being treated with those drugs, but it will help us identify areas of unmet need.

I'm going to move on quickly because I'm running out of time.

The new issues. I think we really need regulatory guidance for pharmacogenomic development. These are coming I think from the agency, and basically I think the key issues are going to be how do we co-develop diagnostics and drugs or biologics and basically coordinate the regulation and delivery of those compounds to the market, and those tests. We're going to need new technology to deliver these tests. Discovery tools will evolve, but the key issue is going to be developing and standardizing multiple analyte tests, which is something that we don't do effectively in the clinic yet.

I think a key important issue is related to education and the resistance and fear of genetic testing. The most important thing we can do is protect privacy of genetic information, treat it like other medical information, and this is really key. We get patients refusing to join into pharmacogenomic studies because of fear of genetic information and genetic privacy, and I think that clearly is something that is of enormous importance.

Developing clear guidelines, and we're seeing those coming from the FDA later this year. I think an enormous contribution could be the public funding and support for pharmacogenomic research in terms of databases and tissue banks using appropriately anonymized and appropriate informed consent to support studies in these areas. Supported public-private consortia, the SNP consortium, the Mouse Genome Sequencing Consortium are examples. I think the ongoing International Genomics Consortium for Cancer Transcriptional Profiling is another area which is really meritorious of public support.

Again, the continued support of public education about genomic sciences.

I'm sorry if I ran over a few minutes.

DR. McCABE: Thank you very much.

(Applause.)