STEVE: Well, welcome. It's really a pleasure and an honor to introduce the speakers today. And I want to do it in two ways, I want to introduce them in a minute for their detailed credentials, which are superb. But I want to bring us back to this concept that's on the slides. So we had a chance to have Chris Austin, head of NCATS, give us a nice talk a few weeks ago on what is arguably one of the most major hurdles in health care and health care costs.

And that's summarized by the Eroom's Law slide here. Where we're looking at not at a constant increase in cost for any therapeutics, doubling every 9 years. As bad as that is, what's really horrific to all of us is when that graph goes to zero, which is 2040.

And so we need to come up with new and innovative ways for therapeutics. And this is the broad vision, and so the next slide is, where do we get our current drugs from, for therapeutics. And this is a slide summarizing it and I wanted you to be thinking about it is from the sources, so synthetic source. We heard a talk a couple of weeks ago on Sanford Burnham and this wonderful partnership for innovative science for synthetic chemicals. And

But that's really only 1/3 of the pie or maybe half of the pie. The other class, right? Are actually natural products. And many of our frontline drugs are from natural products. And there are lots of advantages of natural products. They often are scalable, so they're often very cost effective, which is important not just for a rich Western economy for health care, but we can afford a drug that $2,000 a month in principle, but for be able to scale for around the world. But also the fact that many of these natural products are put in people every day, and haven't done so for hundreds or thousands of years.

And a lot of the components have some sort of safety. That's another part, so entry point, but what's often been a challenge has been sourcing, finding a good high quality, reproducible source of these products if we want to go ahead and use them as a source for novel therapeutics. And so my introduction and background to this, was when we were doing some work in this moving beyond repurposing drugs, and we had a chance, an opportunity to connect up with the speakers today. And we've actually put in one initial application for a P50 grant for a center for the Mayo and Iowa State Center for Resilience. That's the space we're going into.

But as a part of that, it became really obvious that the fantastic science they bring, could be applicable to many areas here at Mayo. So I wanted to have them come down, I wanted to introduce to this audience, which is really the audience of looking hard at how are we going to innovate, how are we going to come up with and attacking Eroom's Law, and look at it from a new point of view. And so with that, that's the broad vision. So I hope that not only does you get a good feel for understanding of the interesting science that they have, but I'm happy to help facilitate. I'm sure the CCATS would help facilitate a potential collaborations in this space in the future. I hope that's helpful.

OK, so our two speakers today, Eve Wurtele, professor of the genetics department, genetics developmental cell biology, I think GDCB at Iowa State. She has a terrific background from UC Santa Cruz, and a PhD from UCLA. She did a post-doc in UC Davis, and was a senior scientist at NPI, Native Plants biotech company, before she joined Iowa State. And she in 2012, received the Pioneer DuPont award for Iowa Women of Research Innovation and Leadership. She brings her plant genetics and bioinformatics skills really forefront today.

And then Basil Nikolau, is the Frances M. Craig professor in biochemistry, molecular biology at Iowa State. Director of the Center of Metabolic Biology and director of the Keck Metabolomics Research Laboratory at Iowa State. And brings wonderful biochemistry to the table. And in plants, and they really represent, to us, this huge access to the science of plants as botanicals. And we're excited about the possibility of this becoming a major collaboration for new therapeutics. So welcome. It's a real pleasure.

BASIL NIKOLAU: Thank you very much for the invitation, for the very kind words in introducing us. We've never done this before. We've been working together for almost 30 years now and never team tagged so let's see how this happens. Hopefully everything is going to work out and no repercussions later tonight.

OK so, as Steve said I'm a biochemist, I studied plants particularly. I came into this world in terms of plant biochemistry, primarily because my PhD advisor was working in plants. And particular emphasis was on lipid metabolism in plants. And as the years have gone on, as you get older, maybe you hopefully wiser, it's gotten wider in terms of the emphasis and sort of trying to understand metabolism in a broader sense. And so that's what, and had never gotten more and more interested in what I'll talk to you about in a little bit, about what is called specialized metabolism, or seigniory metabolism is the older term.

OK so first the disclosure, no conflicts of interest to disclose or anything. So let me move on. So what I'm going to talk to you about is this concept of natural product chemistry or biochemistry. That has a very old tradition, it goes, it's actually how biology became biochemistry. When chemists got interested in trying to understand natural systems. And then this new word defined metabolomics, as to what is metabolomics and how it helps decipher metabolism. And how we conduct metabolomic studies and integrate it into the wider concept of genomics, as the genome is expressed. That's my charge here in the first 20, 25 minutes or so.

And then Eve will follow on and, I don't actually know what these words mean, but the long battle of standing still. Oh I know, that's evolution and the survival in a plant standing still. And then new resources that we have been developing and how to integrate metabolomics and genomics together.

OK so natural product chemistry, as I said to you, is an old, has a long history. It's how we, a lot of the sort of common names in lipids where I came from, come from this world where fatty acids were first identified and in various unique places alactacid, for example, was first identified in olive oil, and that's the name. And there are a number of common fatty acids like this that have common names associated with the source from which they were first identified. But basically, the initial drug designs that came into play, and I'll show you a few slides of these that are famous now, came from this analysis of natural products, of natural systems, particularly plants.

And these were first they were just random, chemists just having fun, but later was bio activity driven purification. But as the years have gone on, and of genomics have come into play, it's become a much more rich environment to play in. There are a lot more resources that are available. And so you can now look at this problem from a perspective, of not just chemistry and understanding the chemistry of the structures that you're identifying, but also integrating genomics where we have capabilities of entirely sequencing the blueprint of an organism. And that's integrating that information with chemistry and have a better predictive capability in terms of understanding how these small molecules biosynthesized. And and using the promiscuity of enzymes to create new, non-native natural products that could have activity.

OK so famous examples are on the slide. And I go back to aspirin to the 1700s, which was isolated from the back of the willow tree. Penicillin , is a very famous example, in the early 1900s lovastatin is a fungal product that is based upon a product that was identified in the 1970s. And more recently taxol, which is from the back of a tree in the Pacific West coast. And taxol is now being synthesized as a hybrid biologically and chemical conversion processes.

So this is the history upon which we sort of are jumping into. And I'm going to tell you a small story about from echinacea, which we have been studying for the last 4 or 5 years. And it comes out of Native American Indian folklore. And they use it via use of the roots of these plants for medicinal purposes.

OK so first, I just want to do a few definitions. In terms of metabolomics, as the word implies, this it's a fusion of the word metabolism and genomics. And it means a couple of things. Functionally, experimentally, what we're trying to do is just like a genome. You start at one end and you sequence to go to the other end. The idea is to comprehensively catalog all small molecules, metabolites, that an organism has at any one time, at any specific tissue type.

That's one way to look at it. But the other way that I'd like to look at it, is in terms of trying to understand metabolism in the context of the entire genome. And that's a little bit of a mindset, rather than a mechanical thing that you do. The first bullet is about how mechanically you determine the metabolome, but the second, is how you decipher information from the metabolome and integrated with genomics to create models as to how these small molecules are biosynthesized.

And this is where you really need to, and I'll show you, I hope I can show you, is that you need to understand chemistry. Because chemistry in itself has predictive language. And that's the language that enables you then to integrate genomics and make predictive models as to how biosynthesis and metabolism is occurring.

All right so, so metabolism, I think most of you are aware, is that the enzyme catalyzed reactions that all biology conducts. It's the way we capture and energy and the way we build structures biologically. We categorize it into two, at least in the plant biochemistry metabolism area, we tend to categorize it into two broad categories. Primary metabolism is everywhere. It's the same everywhere. It's the boring stuff, it's the stuff that you learn undergraduate biochemistry textbooks, glycolysis, the TCA cycle, amino acid biosynthesis, that's common to every organism almost. It's a little bit different here and there, but it's pretty well universal.

Secondary metabolism, or specialized metabolism, not to say it is secondary in importance, but the specialized metabolism is specific to clades, to taxonomic clades. And it's what I like to call the spice of life. It's what makes cinnamon, cinnamon. And bailiff, bailiff. OK? It's that unique capability that only a few clades, taxonomic cldes, have capability of synthesizing. And it is these molecules that give rise to most of the drugs, actually, that we've been used. So the second slide, third slide that I showed you, are almost all secondary metabolites, so a specialized metabolites.

So you tend to, this figure, you can buy this from Sigma, its primary metabolism. There are approximately 6,000 metabolites in this figure. It's the common part of metabolism that is present in all organisms. And it's gene encoded as a secondary metabolism. But it is as I said, it's common to every biological organism. The idea is how do you integrate then genomics into this map? And decipher the metabolic processes, or the enzymes, the genes, and then the biochemistry that gives rise to secondary metabolism.

So technologically we become very, very good and it's become relatively easy to sequence the genome, where the blueprint is at. And genetic information flows down, as the arrow indicates, the blue arrow indicates. We have the rules of how to connect the genome to the transcriptome, then proteome, that's the central dogma. But going from the proteome into metabolic flux, metabolic conversion. And that's the definition of metabolome, and then ultimately the phenotype of a particular organism, that's the hard part. And that's the part that actually needs chemistry. You need to understand chemistry in order to understand the connection between the enzyme and the metabolome. The protein that catalyzes the reaction because within that chemistry, within that knowledge, within that structure of the metabolite there's information as to how it's put together. OK? And I'll show you that in a minute.

So it's increasingly difficult technologically because of one primary reason. The genome, the transcriptome, if you think of these molecules as polymers, the genome and the transcriptome is a polymer of four monomers. Proteins are polymers of 20 monomers. But the metabolome, is a monomer of individual elements. And the complexity, and how it's put together, is a little bit difficult and to understand. And it's not easy to translate that information from the protein to the metabolome because that activity, that catalytic activity, is dependent upon the three dimensional structure of the protein. And that's difficult to a priori predict.

What we have done in the past 20 years, as we've become better and better at sequencing the genome, is that we've looked at this problem from the top down. Prior to having genomic information, we would look at it from the bottom up. We would look at some organism, we would look at some biological system, where we would chemically analyze it and then try to work upwards, back up. But in the last 20 years, I would say, we've become increasingly top down.

And I think that we need to go back a little bit and try and go top down, and bottom up, and try and meet in the middle. OK? All right so, the challenge of metabolomics, I sort of alluded to this, is that we don't know when we finish. OK? So if people have made an effort to understand how big is the metabolome of an organism. And it's very difficult to know that, to count every molecule, small molecule that's in an organism. People have estimated that biological systems have the capacity to synthesize something like 200,000 different small molecules, small molecules meaning less than 2,000 daltons. But nobody's actually gone out there and measured them, found them.

And the other challenge is that genomics, as I alluded to being a polymer of four monomers, the technology for deciphering that sequence is relatively simple. But in terms of metabolites and small molecules, it's a lot harder and each molecule has different physical chemical properties. So its isolation is unique to that molecule. And yet we would like to obtain information about the entire metabolome all at once.

So that makes it a little bit difficult because analytical capabilities are targeted to a specific class of molecules. So you tend to isolate molecules that are miscible in organic solvent, and that's your lipids. And if they're miscible in water those are your polar molecules. And that's based upon the elements that constitute each of those molecules, and it's whether they're water soluble or not. It all depends on how much oxygen and how much nitrogen is in those molecules.

So to measure the metabolome and determine how big it is, and analyze the entire metabolome, you need a large number of different analytical tools. And so that's one of the things that we did at Iowa State about 10 years ago. We were lucky enough to get an endowment from the W.M. Keck Foundation, and we built this metabolomics laboratory, which basically has about eight different analytical tools in it now and machines. And we can separate and analyze metabolites based upon chromatography of some kind, gas chromatography, GC, liquid chromatography, LC. And then this instrument at the bottom, Fourier-transform mass spectrometer, which weighs molecules. This mess spectrometer weighs molecules to very, very high precision, to less than one part per million.

And that enables you to easily determine the elemental formula for each of the metabolites. And that's very, very important because as I said to you, we need to know the chemistry, the structure of each individual molecules, because that's the information that enables us to tie the metabolome to the proteome. So when a customer comes to us and says we would like to analyze something, we work with them and actually, the graduate students, it's mostly graduate students, post-docs that come to the lab, we let them have access to these instruments, and they analyze their own data.

So that is a teaching tool because most of our customers that come are biologists, and lack this chemical capability, this chemical knowledge. So by forcing them to work on the instrument, they actually gain knowledge of chemistry, that's one. The other thing that we found is that if you give your sample to a person and say, here analyze it for me and then I'll come back in a week, the person that's analyzing the sample for you isn't really vested in the analysis. Whereas if it's a graduate student, he/she is vested in that analysis and they will do whatever is needed to get the results that they need. And that's important.

OK so this is the flow of the analysis that we do. We start with biological material on the left and we make extracts and it's important how do you make those extracts of the metabolites. You need to stabilize the [INAUDIBLE], some of the metabolites are chemically unstable. So you need to be careful how you do that because you want to capture the metabolites as they are in vivo, without any modifications that you introduce as you extract them. And then you get them ready for analysis. You may do some chemical derivatization to enable that.

And then you go through separation. And separation is primarily gas chromatography, where you put them into the gas phase and separate them, or liquid chromatography. And then after separation, you have to detect them. And there are various different ways of detecting them. As you may have alluded from the last slide, primarily we use mass spectrometry. And mass spectrometry, many different flavors of mass spectrometry, but basically it weighs molecules.

And the way they are weighed, is in order to weigh them you have to charge them, you have to give them charge. And so either you were bombarded with electrons or you treat them with a chemical that gives it charge. And that charge enables, that charge molecule through a magnetic field, and the way it goes through the magnetic field depends on its mass. And that basically gives you the information about the weight of the molecule.

But some of these, depending upon the amount of energy you put into that ionization process, it also fragments the molecule. And when you fragment the molecule, it gives you information about structure. And that then is a jigsaw puzzle, basically. So imagine that's the molecule, you bombarded it with an electron, it fragments each piece, gains one charge, but you not only get the individual pieces of the individual pieces. I can point here somewhere. OK and then not only do you get the individual pieces of fragments, but you get the various combinations. You may get two pink ones together, a pink and a blue one together, three pieces together, four pieces together, five pieces together, six pieces and it's a matter of then trying to put that puzzle back together again, based upon the weights of each of those combinations.

So that's the sort of the mechanics of how we do metabolomics and try and decipher the information as to the thousands of molecules that are present in a particular tissue. Using all the instruments that I showed you, by combining data from all of those, we work with plant systems or microbial systems sometimes, we can obtain approximately 2,000 individual metabolites from any one particular sample. You have to work at it, but somewhere between 1,000 to 2,000. Of those, about a third we can definitely chemically identify. We can name them this is molecule X, this is molecule Y, this is molecule Z. 2/3 are unknowns. And we've never seen them before, or if we've seen them before, we haven't worked out the chemical structure.

And knowing as I said to you, chemical structure is important. And so that's when isolation is to occur and then analysis to obtain chemical structure is now primarily done with NMR. And we have that capability next door to the metabolomics lab.

All right so, let me tell you a small story about echinacea. This was a project that we started about 10 years ago. I got interested in it because of those kinds of molecules that are on the right there, which are called alkamides. They're very, very unusual molecules that don't occur in many different places. For those of you who are a little bit of a chemist there, they are amides. And so there's a nitrogen linked to a carbon that has an oxygen on it. OK.

BASIL NIKOLAU: So The side of the molecule, so the idea is to try and understand what metabolic process, taking in plant systems taking CO2 and ammonia from the soil and some solar energy, etc. What flow of information, what flow of metabolites, which genes, enzymes, does the system use to make those kinds of molecules. OK so, there's a nitrogen containing portion of it.

And there's the other side, on the left hand side, is a fatty acid containing molecule. The interesting part is that these fatty acids are decorated with carbon carbon double bonds, and carbon carbon triple bonds. And those are very rare, particularly double bonds that are next, not quite next to each other, but there's a double bond, single bond, double bond. And those are very, very rare in biological system. And triple bonds are even rarer. And so how do these molecules, how these molecules are assembled.

OK as I said to you, there is a medicinal history to this, Native Americans use these roots of these plants. And there as nine, as the slide says, there are nine different species throughout here, actually in the prairies you may have this in your gardens in the summer time. And across the different species, the nine different species, approximately 30, 40 of these kinds of molecules have been identified.

So the way the molecule is put together, it's inherent, as I said to you, in the chemistry of the molecule. The right, now I've switched around sorry. The left hand side now, as I've shown you, the blue side is the nitrogen containing molecule. And that comes from amino acids. And if you look at the structure of that blue part of the molecule, it either comes from valine, which is the top the top structure. Or the bottom structure, is isoleucine.

One of the enzymes that we're going to be looking at then, is to take valine and decarboxylate it to make the amide. Or take isoleucine and generate the amide. That molecule that contains only nitrogen would be the amide, the blue part. And then you have to have fatty acids with some unusual biochemistry that has to occur in those fatty acids to make those unusual structures that I showed you before. And then there has to be an enzyme to put those two molecules together.

So we came, that's the simple model. Here's the more complicated model. OK? The top yellow box is the module for providing the blue part of the molecule. The bottom, we're not absolutely sure which one it was, there are three different ways of doing it. And this comes from just knowledge of biochemistry. And one of those three, is a possibilities. And there's some enzyme that we call the ligase that puts these two molecules together.

So we went into echinacea, it made extracts from different parts of the plant, approximately 36 different plants, we sequenced the transcriptome of every one of those, determined the metabolite for every one of those, and then we compared the abundance of the metabolites, and correlated that abundance data, to the transcriptome data. And that's sort shown on the graph on the right.

The fact that these, so the graft on the right, shows you the abundance of metabolites on the x-axis. And different metabolites, different issue types, or parts of the plants, that we analyzed. The 36 different parts that are numbered on the left. And the fact that these lines are parallel to each other, shows you that they correlate in their abundance. And that's the tool that we use. We look for correlations between metabolite abundances and gene expression at that messenger RNA level. And based upon that, we then create models and try and identify enzymes that would catalyze the types of reactions that I showed you here. We need a decarboxylase, we need a ligase, and we need a fatty acid modifying enzyme.

And that led then to finding decarboxylases. Decarboxylases are marked because they carry pyridoxal phosphate as a co-factor. And that's recognized in the primary sequence of the protein. And so we found all the decarboxylases, or the pyridoxal containing proteins that were present and in echinacea, and compare them to other known decarboxylases from other organisms. And found the red dotted one, that I showed you that is matched with a red asterisk, and that turned out to be valine decarboxylases. And the way we prove that it was valine decarboxylases, is that we expressed that protein, that gene in E. Coli, in this case on the right, purified it, and carried an enzymatic studies to show and demonstrate that it is able to get to decarboxylate valine. And it could also decarboxylate isoleucine. So it's one enzyme and it's promiscuous, and can use both substrates, although it prefers valine rather than isoleucine. And that fits the profile of the metabolite that most of the acclimates are valine derived.

OK. Similar strategies, I don't have time to go through all of them, but I'm not going to go through it all, we have found the ligase that it puts the enzyme together. And again, it is the same kind of study, looking for correlations between transcriptomes and metabolite data. And looking for motifs that I shared that define it as a ligase. And we found the ligase, and we also found the fatty acid modifying enzymes that are used to make the double bonds and the triple bonds that I showed you before. And so it's that middle pathway that decorates those fatty acids. So based upon this information then, we've rebuilt the pathway.

And so one of the things that we now are doing, is expressing all these genes that we've isolated. It turns out to be approximately six genes all together. So we're expressing these in a heterologous system and a microbial system. So we can recreate this metabolic process and make these kinds of molecules in the lab, rather than in a planter.

Very quickly, so that's how we sort of use metabolomics to discover new genes, new enzymes. And we reconstitute enzymes and pathways so that we can make these molecules in the lab. And then the challenge then becomes to optimize that system. All right, one of the things that we, one of the things that we has become clear to us, as we've been doing this work, that you need high spatial information about the distribution of metabolites. And so we need to understand, because that enables you then to correlate transcriptome data and metabolome data much, much better.

And we can do that for genomic, for genes with DNA, you can do it for RNA, we have the technology so we can do it for proteins. So we can locate these macromolecules to individual cells and even individual subcellular compartments. But for metabolites, that's a very difficult thing to do. So that's what we've tried, I've said I have a collaboration, a very nice mass spectroscopist in the chemistry department, who I'll acknowledge at the end. Who basically rebuilds mass spectrometers, I mean he builds them from the ground up. He buys them, and then rips up the warranty, and drills in there, and does what he needs to do.

So basically, if you're familiar at all, this is what's called MALDI matrix assisted laser desorption ionization mass spectrometry. OK so basically, the mass spectrometer shines a laser onto a thin microscopic slide of the tissue. And when the laser hits that spot, molecules fly up, and you weigh them. And you get the spectra that's shown on the right. And then you move the section a little bit, you repeat the process, and a different spectra arise. And you are able then to reconstitute the image of the metabolite relative to the xy-coordinates of the slide. So that gives you the image and the distribution of the metabolite on that slide.

OK, so we started to do the same thing and applied it to plant systems. And we started back in 2008, we had a spatial resolution at that stage of about 100 microns capabilities. I'll show you a few slides in a minute, what I told you, he rebuilt the mass spectrometer, he realigned the laser so it was a much finer laser spot. Most commercial instruments shined a laser that's approximately 50 microns in diameter. We're down, I'll show you in a minute, down to five microns.

Anyway, over the years, we got to approximately 10 micron resolution. By controlling the stage and making sure we knew exactly where it is. And this is showing the distribution of a particular metabolite, a linear hydrocarbon. This is a molecule, this is an arabidopsis flower. And if you're not sure, a arabidopsis flower is a couple of millimeters in size. You cannot dissect it, I haven't been able to convince any graduates to dissect the various different parts of the flower apart. So this is a nice way of getting a really high resolution data about metabolites.

So you can get, if you really push the system. this was as I said to you a few years ago. We can see individual cells here, individual pollen grains. And that the molecules that are there and the amount of each of this molecule that's there. And this is the rebuilding of the mass spectrometer. Mostly the re-aligning the laser and re-aligning, that's the red bean, so that it is much, much finer focus so we're now down to five micron spatial resolution. And when you do that, and this is the type of data you get, this is a cross-section of a maze plant, of a maze leaf. Which has five or six different cell types in the cross section and the distribution of the various metabolites are color coded, showing where they are in the leaf, in different cell types that are there.

There's interesting biology associated with this, but I don't have time for it and it's very, very plant specific. I mean, we grow corn all around us. There's a lot of very applications to this knowledge for agriculture. But it just illustrates that sort of the power of having this capability and the applications wide ranging.

All right, I think I'm going to stop there. I just want to acknowledge the students that have participated in this. And that Young-Jin Lee is the professor in chemistry who has been collaborating with, who we've been developing this imaging capabilities. And the others are post-docs and graduate students who participate in the various projects that I showed you. All right.

EVE WURTELE: OK so one of the things about plants, is they're stuck in one place. And if you're a cow or something, you just come along, eat the whole thing, or if you're a person, you pick it, eat it. And that's really tough when you're a plant. They don't have, plants don't have many things they can do about surviving. One thing they can do is create chemicals. And they create chemicals, they've been doing this for 500 million years now. And the environment's really brutal. There's microbes, there's larger predators, and and there's even plants that are growing faster than other plants and shading them, or taking the water from the soil.

So plants are battling against everything and they're stuck in place, and they really do most of their battles by either making chemicals constitutively or producing them upon demand when they sense an attack. And the number of chemicals, biochemicals and metabolites, that are made by plants is getting upwardly revised over time as our machines get more sensitive, we're finding more compounds. And so we're thinking like 200,000 compounds in the plant kingdom is the current estimate and it'll probably keep it going up.

So each set of species, has its own set of metabolites that it uses to cope with its environment. And for the most part, we don't know even what these are, let alone why which compounds is there. But think of how much energy it takes a plant to make compounds. So they're taking primary metabolites, they could be using for growth, and they're making defensive compounds instead. And then they have to make special compartments, usually to keep these compounds in, because if they don't make those compartments, the compounds will affect the metabolism of the cells that make them.

And on top of that, there's a tremendous amount of energy used in terms of DNA, and RNA, protein. So there's a big it's a big energy drain to make biochemicals. And so one of the most advantageous ways, if you're a plant, to fight predators is to make very highly bio-active compounds, and not make too much of them and just sequester them somewhere. And and that's why over all this time, we have all these compounds that had been produced through evolution and that are highly bio-activity.

So Basil talked to you about some plants that have been used medicinally for a while. I want to talk to you about new resource, this is just one example. So there's a whole lot of plants that have been barely studied, this is one of them. This is hypericum gentianoides and we picked this, because it's been used by Native Americans. And there's actually one sentence in one book that says it was used by Native Americans, by Cherokees. That's all that we know about it's use. It's endemic to the US, it's weeds and it grows by the roadside in a lot of the southwest, southeast. And it's related to St. John's Wort, which you probably are familiar with as a very commonly used medicinal plant. And you just go to Hy-Vee, your Wal-Mart, or anything, and there's five or six bottles of St. John's Wort that you can buy. You can't buy hypericum gentianoides. Package

But we were interested in why this might have been used, what's in it. And so nobody knew anything about it at the time. So what would it turns out that the major specialized metabolites that are in this are these di-acylphloroglucinols that you can see up above. And the blue and the red moieties are the acylphloroglucinols and di is basically that they're conjugated together. And there's a number of them in hypericum gentianoides. And you can see that where the R groups are, they vary by their decorations.

So for example, hyperbrasilol C on the R4 group, up at top red ring, has a long acyl chain. So they're decorated in different ways and depending on how they're decorated, actually, it makes a difference in terms of their bio-activity.

So by analyzing the compounds, a whole variety of compounds of putative precursors, we've started to work out a pathway. And working out this pathway of how these are synthesized, you can see these compounds, you don't have to see what they are. But the ones in blue are compounds that nobody's ever discovered before. And so this is in a period of just a couple of years, we identified these new compounds. And no idea if they do something or not, these compounds are acylphloroglucinols and they dimerize to give rise to these di-acylphloroglucinols, of which some of them are in other hypericum species, and others aren't. The [INAUDIBLE] and blue are not.

So we have already these 30 or so compounds that are present in relatively high levels. And if you look where that arrows point, there is a gland, and you can see down below the white arrow, that's a cleared leaf and you can see the gland a little bit better. So leaves have these transparent glands and if we used the laser absorption, which you just saw in Basil's presentation, and we monitor for the particular compounds that we know are present. That are there are natural products in hypericum gentianoides. You can see that they're all located in the same region. Each compound is monitored separately, but it's the same leaf slice. And basically what you see is they're superimposed exactly on the glass. So these glands are made, and they're specialized, to make these compounds and we also can look at all the precursor compounds that are also present in the glands. So basically these glands are little factories to make these natural products.

So do they do anything? Well, turns out that they do. So our group, and several other groups, have found that they alter calcium homeostasis, they have anti-inflammatory activity, they're anti-depressive in some situations, and have anti HIV activity. And there haven't been a lot of studies, but the studies are quite consistently that they have bio-activities.

Now, we can look at the whole extract, or we can look at the individual compounds, or fractions thereof. And in a give bi-activity assay, different compounds are more active, or less active, or not active at all. So we went on, in this case, to take the HEK cells, which had altered homeostasis. And we treat them, and then do RNA seek. And look to see the difference between mock treatment and treatments with extract or compounds from h. gentianoides. And essentially what we saw, is that there's a lot of activity changes in activity of key genes in anti-inflammatory pathways. And the red circles are things that are up regulated and the blue circles are things that are down regulated. And we find that two compounds and the extra to this.

So now we have new compounds that have bio-activity and they really haven't been studied much, but there's a tremendous space for them to be studied, because they're doing something. The plants evolved them for a purpose and we can maybe milk them for all they're worth, for our purposes. OK so, a lot of people are now studying natural products, and not necessarily just not for products, but any kinks of pharmaceutical actions. And what's the last part of what I want to talk to you about, is how important it is to keep this in a database, so that everybody has access to it.

So I just might illustrate that here. So for plants and microbial species, other species that might have bio-active compounds, we can look at what are the metabolites, what are the genes, where they are located. We can select plant materials and from that we can gather data and put it into a database. And the same, we could also take those compounds do bio-activity guided fractionation, determine which fraction's are active and which are not active, which compounds are active and not active, in an animal model of some sort. And then we can then do the same thing with that animal model, we can do metabolomics on that transcript, metabolomics on that, put that in a database, as well. As the database, if the database is a sophisticated enough, it can start to give us information about what are the compounds that are bio-active, and what genes are they affected.

Now we're not there yet. But we do have a database, it's a good start. So this is called PMR, plant eukaryotic and microbial systems resource. And Man Houy is the guy on the right, and he's smiling in the snow there. But he's the main developer of this resources, he's a fantastic computer programmer. And what we have in this resource now, is data from a number of species, and experiments from the species. We have now 25 species and about 14 terabytes of data. And what PMR does now, is it stores and retrieves metabolomics and transcriptomics data, and it's integrated with other databases.

So I'll just quickly show you a couple things. Most done here. So it does different kinds of linked statistical analyzes. So this is called the volcano plot, and basically, each point is a different compound. And this shows the differences, you're comparing two samples here, on the left is ripe fruit, on the right is sterile seedling. And you can then, these are compounds that are different between the two, and you can click on a compound. Look at its pattern across different samples, you can analyze the genes that are correlated with this, and you can get q values and p values for this, you can get the Pearson's correlation. And you can then look at the genes and you can look at the genes, for example, one gene that you think is correlated with metabolite.

Here's an example. This is a gene that's correlated with hyperforeign, and we're thinking that it may likely be isobutyrophenone synthase, which is the enzyme, which is the first step in making hyperforeign, and other acylphloroglucinols.

You can look at the domains and overrepresented pathways, and basically it's a community database. And it's set up for anyone to explore the data and develop their own hypotheses. So if I put my data into there, and as soon as I published on it, it's released. Then anyone can go in and look at that data, and use the statistical tools and analyze the data, for maybe they're interested in a different pathway, a different compound. So they can apply their interest to that data. They can also download it and analyze it in any way they want.

And so this is to me, it's really important to enable the community to work together to best solve these kinds of problems. So any researcher can deposit data. I'm not sure how many people we have, we have 40 or 50 people that have deposited data so far. And you're welcome to do so as well.

So thank you. These are the people that have worked, and I want to also particularly thank Steve and also Carl and Margo and Jeff and Eric and Mara, who are the Zebrafish, our Zebrafish colleagues that we've been working with lately. And our support as well.