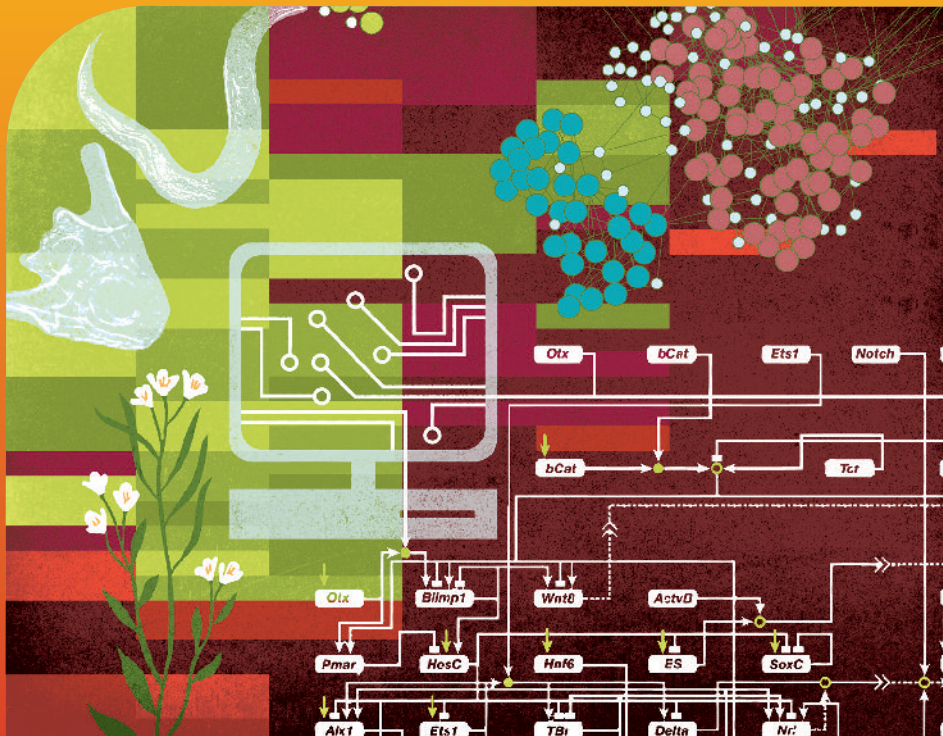


Genome LIFE



It's the Network

In Genomic Age, Systems Biology Approach Takes Root in the Search for Fundamental Insights

In the old days of molecular biology, scientists tended to focus in on genes one at a time. Those genes were assembled into linear pathways, each of which took several years of work to put together. Then genomics came along and changed everything. Suddenly, labs could generate not just linear paths, but complete parts lists. Systems biology was the natural next step and, in systems biology, the network is king.

"Systems biology realizes that the key part of biology is the connections between pieces," says Philip Benfey, Director of the IGSP Center for Systems Biology. "It's not the pieces themselves but their relationships to one another and how those networks operate dynamically in varying conditions, and in different organisms, across different time scales."

That kind of global approach requires close collaborations between experimentalists and theorists – those who are generating enormous genomic datasets in the laboratory and those with the quantitative skills to distill that complexity down and devise ways to represent it meaningfully – within a broader culture of science that is supportive of such interdisciplinary exchange.

"The human mind is very bad about grasping temporal relationships and making sense of them when more than two or three objects are involved," Benfey said. "You need a mathematical formula to reduce the dimensionality and figure out what is important. Out of that, discoveries emerge that could never have been guessed at by staring at the data or a list of parts."

Systems biology had already begun to take hold in the early days of the IGSP, with efforts by Joe Nevins, Mike West and a young Erich Huang to gain a more detailed understanding of cancer-causing gene pathways and the interconnections among them (see "Interlocking Pieces," p. 7). Systems biology continued to grow organically out of research collaborations between faculty representing departments in biology, statistics,

Systems Biology (continued on pg 3) >

INSIDE

Meet the faculty and students who are using genomic tools and systems biology approaches to find out how hungry worms pause their genomes, how embryonic mice and red-eared slider turtles choose their sex, how microbes survive in some of the saltiest places on earth, how seemingly conserved developmental gene networks can evolve over time, and much more.

Message from the Director

When Duke proposed establishing a new campus-wide Institute for Genome Sciences & Policy, it was conceived as a highly interactive group of interdisciplinary centers involving faculty and students from across the campus. This model has served both the IGSP and Duke well for the past eight years, producing a series of thematic but flexibly run centers to coordinate development and implementation of IGSP missions in research, education and service. Over time, individual centers have adopted different “personalities,” and the roster of IGSP centers has adapted to changes in the genome landscape, both at Duke and beyond. Natural clusters of faculty and students have sprung up around specific areas of investigation – sometimes within centers, sometimes between centers and sometimes under the broad IGSP umbrella itself.

Systems biology, featured in this special issue of *GenomeLIFE*, is a perfect example of this theme-based and program-driven interdisciplinary emphasis. Driven by an outpouring of genomic data and the need to fit the pieces together, an emphasis on networks in biology and medicine grew out of groundbreaking work by Joe Nevins, Mike West and their colleagues in Medicine and Arts & Sciences over six years ago. As the field of systems biology emerged nationally, a broad effort took root out of natural collaborations between experimentalists and modeling experts from biology to engineering, medicine to statistics. Those efforts, fueled in particular by Philip Benfey, coalesced under the IGSP Center for Systems Biology, a structure that has enabled critical interactions and connections between people, sparking new ideas and research directions and fostering the interdisciplinarity at the core of University Institutes at Duke. As you will see, systems biologists of all stripes are now exploring the intricate biological networks that govern living cells in all kinds of organisms and on all kinds of time scales. It is this kind of success that I imagine Duke’s leaders had in mind when the IGSP was conceived ten years ago.

But the world of genomics – whether viewed through the lens of science or society – is unquestionably very different now than it was back in 2000, or even just a few years ago when the Center for Systems Biology and other centers launched. While the demands of anticipating and adapting to a landscape of ever-changing science have shifted, the consequences for

“The world of genomics – whether viewed through the lens of science or society – is unquestionably very different now than it was back in 2000, or even just a few years ago.”

and the engagement of society have only deepened. This invites consideration of new strategies to meet these new opportunities and challenges. Thus, as successful as the circa 2000 model has been, it is time to consider ways in which the organizational structure of the IGSP might adapt to better serve our diverse IGSP missions and to encourage leadership and entrepreneurship in selected domains of science and/or policy.

As we approach our tenth-year anniversary and to assess whether our current organizational structure and intellectual balance is optimal for the future, we have initiated a two-phase evaluation and planning process. Phase I, now underway, involves a review of critical aspects of the IGSP’s mission and structure from an organizational point of view, in light of the rapidly evolving landscape of scientific and social opportunities and challenges that mark our field. Phase II will focus strategically on programmatic initiatives and opportunities, partnerships, and priorities that will guide the shape of IGSP in 2012 and beyond – what one might call “IGSP 2.0”.

In Phase I of this process, a series of working committees are addressing key issues for our future, including our organizational structure, the analytical and service infrastructure, translation and application of genomics, education and training across the campus, and IGSP-led administrative services and functions. Each of the five working committees is charged with performing an in-depth evaluation of the area in question, considering various models for addressing the questions raised, and making recommendations to me and to an overarching “IGSP 2.0” Planning Committee that will guide this process over the course of the year. We welcome and actively encourage input from everyone in the IGSP family, as well as from our colleagues across the Duke campus. Information on these working committees and their progress can be found on the IGSP web site, and comments and suggestions can be directed to igsp2.0planning@duke.edu.

Building on the results of Phase I, the purpose of Phase II will be to make decisions about the recommendations from the working committees and to use that as a base for considering themes, both new and old, to guide the development of “IGSP 2.0” in a way that exemplifies our interdisciplinary and cross-campus reach, spanning both genome sciences and policy.

By articulating our new organizational principles – perhaps quite different from the ones we began with nine years ago – and ensuring that the plan has support and energy throughout the Institute, we will be in the strongest possible position to continue to achieve distinction and leadership in this field. ▶

Huntington F. Willard, Director

Systems Biology *(continued)*

engineering, math and physics. In 2007, with funding from the National Institutes of Health, the IGSP Center for Systems Biology was officially launched and today those early collaborations and others that have developed since are bearing fruit.

“We’ve seen over the last few years that what comes out of a systems biology approach are fundamental biological insights,” Benfey says. “That’s the promise of systems biology, and we’re beginning to deliver on that promise.”

In the pages that follow, we’ll offer a glimpse into some of those discoveries and the collaborations that have made them possible. On page 10, you’ll learn about the varied insights made in Benfey’s own lab based on studies in the model plant *Arabidopsis* and reported just this year. You’ll also

“**Systems biology realizes that the key part of biology is the connections between pieces. It’s not the pieces themselves, but their relationships to one another.**”

—Philip Benfey

meet researchers working on everything from microbes that live in some of the saltiest places on earth to sea urchin larva, nematode worms to mice and even red-eared slider turtles, to address questions about how cells and organisms deal with stress, change their fates and decide which sex they will be. In each and every case, the ultimate goal is the same: to get a little closer to understanding life in all its wonder and complexity.

Systems biology researchers at Duke are also leading the way in some new directions for the field. They are beginning to look at populations of cells or organisms not as amorphous groups but as collections of individuals. That means considering not just how a network behaves on average, but how it behaves, and behaves differently, in one individual compared to the next.

“Stochasticity is fundamentally a part of biology,” Benfey says. “We know in physics we can’t precisely predict where a particle is in space, only the probability that it will be there. We are beginning to see a glimmer of a similar view of biology.”

That heterogeneity isn’t something to gloss over; it is fundamental to the way organisms evolve in some cases and resist change in others. And as technologies improve each and every year and new methods of analysis are developed, the picture will surely get clearer, if more complicated. ▶

Take Pause

If you starve a nematode worm, it will live as if in suspended animation. All growth and development ceases. That stalled activity at the level of the whole organism can be seen at the level of the genome too. The enzymes that copy DNA into RNA literally stop right where they are. Life takes pause.

“It’s as if the polymerases are preloaded in anticipation of feeding,” says Ryan Baugh, an Assistant Professor of Biology. “They are poised for a rapid response.”

Once thought to be the exception to the rule, scientists are increasingly realizing that this pausing mechanism is a more general layer of regulatory control in many organisms and its benefits might go beyond speedy responses. Paused genes could be a key reminder for a dormant cell of who it is supposed to be. That way, when conditions improve, the cell can pick up where it left off without missing a beat.

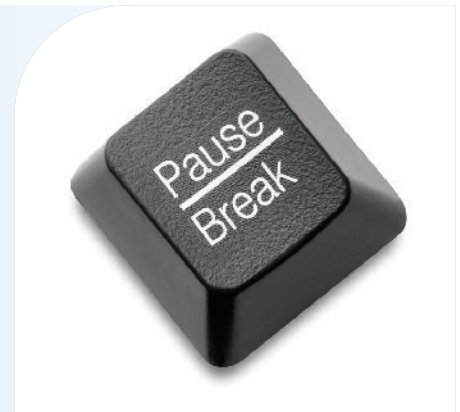
“If you have these developmental control factors bound in the genome but paused, in a sense the cell has a memory of its identity that is very stable,” Baugh says.

Genome pausing speaks to an important property of gene networks and one that is an increasing focus for systems biologists: how networks change over time. “Pausing is a crucial mechanism that can account for dynamics in gene expression,” he says. “It offers a unique way for cells to go from one state to another.”

Ultimately for Baugh, systems biology comes down to a mindset. “Rather than one gene, one process, one cell type, we try to consider several at once and what kinds of relationships they have to each other — using a combination of modeling and experiment to advance beyond what you might get with intuition alone. That’s where systems biology really does offer something new and different.” ▶

“**Rather than one gene, one process, one cell type, we try to consider several at once and what kinds of relationships they have to each other — using a combination of modeling and experiment to advance beyond what you might get with intuition alone.**”

—Ryan Baugh



*When tiny, soil- or lab-dwelling *Caenorhabditis elegans* nematodes go hungry, about 15 percent of the genes in their genomes get set to pause.*

An Alternate Fate

When embryonic sea urchins have just 16 cells, four of them are destined to produce the skeleton. Developmental biologists including Duke's Dave McClay know an awful lot about how those embryonic cells reach their intended fate. But years ago a post-doc in McClay's lab uncovered a surprising twist. He devised a way to eliminate those four skeletal precursor cells only to find that the young urchins stubbornly produced a skeleton anyway. Using new genomic tools and sophisticated modeling, the McClay lab is still trying to figure out just how they do it.

"We know a lot about the gene network it takes to build this one fate," says Dave McClay. "We have no idea how a different network trans-fates to replace it."

McClay has joined forces with Joshua Socolar, a Physics Professor and Associate Director for the IGSP Center for Systems Biology, along with Computational Biology Graduate Student Xianrui Cheng to help sort out the problem using a complementary blend of experimentation and computational modeling. Chen represents a new breed of scientist in the making; he spends half of his time at the McClay lab bench running experiments on the urchins and the other half with Socolar sorting out how to represent the biological phenomena he has observed firsthand in mathematical terms. With both sets of tools at his disposal, he is able to "fill in the cracks between the two fields."

One of the first steps in forging the collaboration was sorting out the language, Socolar says. From a biological perspective, the urchin cells switch from one network to another in the process of changing their fate. But, "from a mathematical point of view, not really. Both have the same DNA and are in some sense running the same program. What's different is that the network is in a different dynamical state. By framing the question in this way, we want to understand how the complete set of network instructions guides the state a cell is in. We want to make progress in understanding the logic of the system."

Socolar says he never dreamed the title biologist would ever be associated with his name in any way. For him, the sea urchin conundrum is inspiring because it brings up fundamental questions that have been circulating in the physics and mathematical modeling community for years. "I've been interested for a long time in systems that undergo interesting transitions in behavior or exhibit very complex behavior," he says.

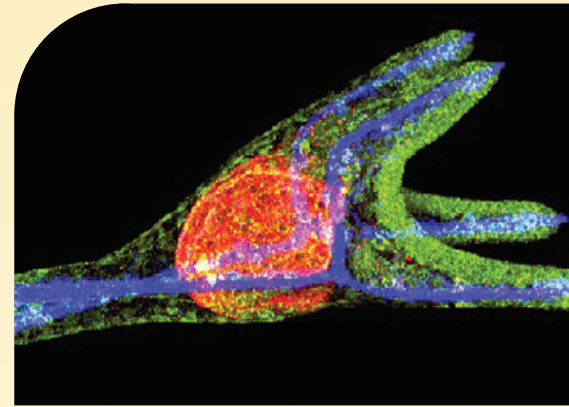
In the 90s, researchers showed that even random networks, in which things were wired together "willy-nilly" could generate intriguing behaviors. That led to a flurry of activity, trying to work out the fundamental properties of a network required to produce characteristics like stability and flexibility. Socolar's interest now is to apply that kind of thinking to address the deeper principles underlying biological networks in nature.

"These biological networks are guided by natural selection and evolution to do things that look very specific," he says. "But, how hard is it really? Is it possible you can get those properties essentially for free if you make a complicated enough network and then all natural selection has to do is just tune the dials to make it more efficient? The random models were interesting in opening our eyes to the possibilities, but now what is really going on in these systems?"

For McClay, it is all about figuring out how cells work, using the new tools of systems biology to address long-standing questions that had just been too big to tackle before.

"All along we knew there was a system, but technically we just couldn't get at it," McClay says. "It was too large a problem. We were missing the genomes and the high-throughput technologies needed to look at dynamic changes in gene expression. Now, all that is available, so what we knew to exist we can actually address. To me, the challenge is to pick the right problems."

He believes what they learn from the sea urchins will have relevance to other organisms and to the kinds of fundamental questions that



A sea urchin larva with its 64 skeletal cells in blue.

preoccupy stem cell biologists of all sorts. They've already made progress that McClay says has shattered whatever preconceived notions he might have had about how the urchin cells operate. It isn't that the loss of the would-be skeletal cells lends some other cell the capacity to trans-fate. Those cells probably have that all along, but the skeletal precursor cells actively prevent other cells from becoming them. "They are the gatekeepers," he says. ▶

“All along we knew there was a system, but technically we just couldn't get at it. It was too large a problem. We were missing the genomes and the high-throughput technologies needed to look at dynamic changes in gene expression.”

—Dave McClay

Evil Twin

The genetic circuits that keep us to a roughly 24-hour circadian clock are no doubt complex. But IGSP Investigator Nick Buchler has a hunch that there might be a simple way to devise circuits that generate the kind of ultrasensitive, all-or-none responses needed to make a good clock as well as many other biological networks.

His work has shown those on-off threshold responses can arise from simple interactions between pairs of proteins – one acting as an inhibitor of the other. He now suspects that sort of interaction may play an unexpectedly large role in the evolution of genetic circuits in nature. After all, protein inhibitors can arise rather easily when genes get duplicated in the genome, a relatively frequent occurrence. The extra (and extraneous) copy is then free to pick up mutations that turn it into a defunct competitor of the first. “Your worst competitor is your twin gone bad,” Buchler says.

Buchler is testing this idea about how real biological systems might be put together not by studying circadian clocks in nature, but rather by building them in the laboratory using the simplest of components. It’s a bottom-up approach to systems biology, and one that comes naturally to Buchler as the first jointly appointed junior faculty member in biology and physics.

“By focusing on simple synthetic systems or running experiments in a computer, we might get a glimpse at what the rules are,” he says. “Patterns

“We might get a glimpse at what the rules are.

Patterns could emerge that tell us how these networks might have evolved in nature over millions of years”

—Nick Buchler

could emerge that tell us how these networks might have evolved in nature over millions of years.”

Buchler describes his lab as a truly interdisciplinary place where two worlds exist side-by-side. Students in biochemistry spend their days at the lab bench, putting simple circuits into yeast to see what it might take to direct the evolution of an artificial circadian oscillator. Meanwhile, physics students are testing out the same ideas in the “parallel universe” of a computer. Buchler says there is budding interest in his young lab from students representing biology, computational biology and engineering as well.

“It makes it fun,” Buchler said. “Everyone is coming in with a different strength and weakness.” While the physicists get steeped in the nuts and bolts of biology, many of the biologists and chemists get their first taste of what modeling can and can’t tell you. “My feeling is that they enjoy the challenge.” ▶

Systems Under Stress

Single-celled *Halobacterium salinarum* specializes in stress. Members of the third and least understood domain of life, the tiny bugs are considered extremophiles (literally lovers of the extreme). They are found in some of the saltiest places on earth, including the Great Salt Lake and the salt flats of South San Francisco Bay. Not only must they tolerate incredibly high and varying concentrations of salt, but they also have to withstand strong fluctuations in light, heat, oxygen and nutrients, all in the course of an average day.

To manage those conditions, *H. salinarum* shifts its metabolism between four different energy-generating modes. It’s a scenario that requires constant adjustment of the gene regulatory networks that run their cells.

Amy Schmid, an IGSP member and a new professor in the Department of Biology, wants to understand how the tiny organism swings that at the genomic level. As a post-doc before coming to Duke,

she was part of a team that devised a predictive model to describe how the environment shapes the expression and assembly of *H. salinarum*’s genes into functional biological networks. After some 500 experiments that measured the behavior of all genes in the genome simultaneously, they devised a model describing the interrelationships among an impressive 80 percent or so of the microbe’s genes and key environmental factors. Now, as an independent faculty member, Schmid is drilling deeper into those interactions in search of the critical links from one network to another.

By studying these oddballs of the microbial world, Schmid ultimately expects to learn a lot about life on earth in a broader sense. “Archaea are the least studied domain of life,” she says. “We stand to learn a lot from them about different solutions to living in diverse environments, and the systems approach offers a way to do that quickly.”

Systems biology is defined by three main tenets,



Photo by Chris Benton

Schmid says. It is collaborative, iterative – with plenty of back and forth between theory and experiment – and globally focused.

“The strength of systems biology at Duke is that it really puts these three tenets to work,” she says, guided by players from an array of traditional disciplines, from computer science to biology, physics to engineering. “It’s systems biology in action and, as a new professor, it’s really fun to be a part of it.” ▶

Network Evolution

In the early days of systems biology, the focus was primarily on assembling parts lists and in understanding how those parts – the genes and the proteins they encode – were assembled into networks. But in reality, those networks aren't fixed entities. They shift and change over time.

"You and I don't have the same network," says IGSP Investigator Greg Wray. "Depending on whether you are healthy or sick, you'll have a different network. Asleep or awake, it's a different network – fourteen years old or 50. We need to think about how we infer changes in networks and, from an evolutionary perspective, that is the raw material for natural selection to operate on. We know shockingly little about that."

Of course, evolutionary biologists like Wray have long known that genes interact and that those interactions must influence the freedom with which organisms can respond in the face of environmental change or other selection pressures. The question was how. Most prior work had addressed the issue in somewhat abstract terms, based on statistical associations between traits.

"Systems biologists care more about physical interactions between molecules," Wray said. "They don't think about correlational structure and the evolvability of traits. The question was whether we could bridge that gap."

To give it a shot, Wray and graduate student David Garfield have turned to a well-studied network of about 100 genes that underlies the development of the larval skeleton of purple sea urchins. They surveyed the activity of those genes in members of a single urchin population including individuals related to one another in precise ways that allowed them to determine whether those differences in expression could be passed on from one urchin generation to the next.

Ultimately, Wray and Garfield hoped to address some rather big questions. "We wanted to think about development as an unfolding process," Wray explains. "When you change something in morphology, where are the tuning dials? Are they right at the end of the process or all the way back at the beginning or somewhere in between? Where can selection tweak that fabrication process?"

The results turned up a lot more variation in developmental networks than expected. "We found all kinds of weird variation in the expression of genes in development, especially early," Garfield said.

That was a particular surprise because scientists had generally supposed that variation in early development would lead to ill consequences for organisms. It appears that isn't so because those early genes have more switch-like interactions. As long as a change in gene expression doesn't flip the switch so to speak, its ultimate consequences for the developing embryo are nil. In contrast, genes that act later in development tend to follow a more linear pattern. Higher activity of one gene more often does change everything else that happens later.

The finding shows that the relationship amongst developmental genes is often very loose, Wray says.

"Just because one kinase is at a higher level doesn't mean its target must be dialed up or down," he says. "There is elasticity in the system. It's not brittle."

“**In the big picture, it's important to understand how organisms can go from a conserved developmental program to a sudden shift.**”

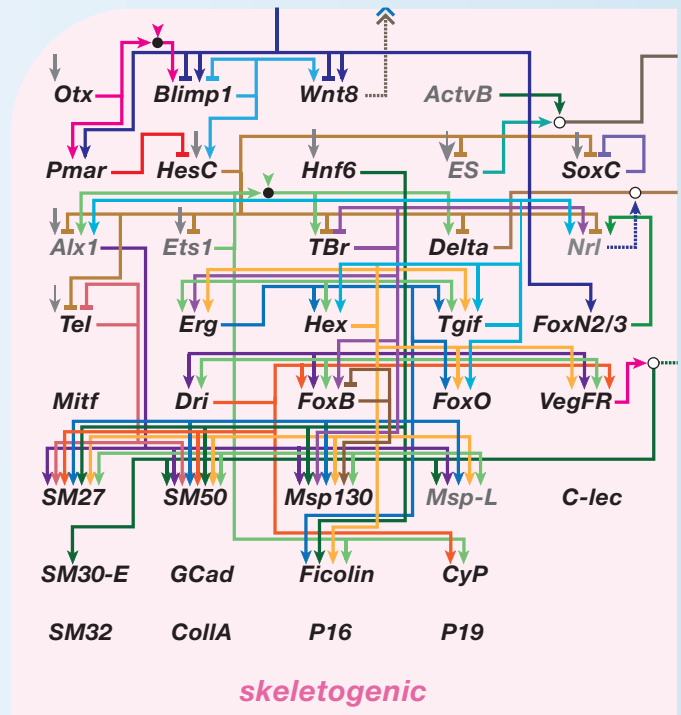
—David Garfield

It's counterintuitive, he says, but in some sense, it has to work that way. "If networks were brittle and highly engineered, then you throw the organism out in the wild and let mutations happen, they would just fall apart. These systems have to have bounciness."

That also means that organisms could harbor a lot of genetic variation that isn't apparent just by looking at them, a finding that may be good news for natural populations in the face of climate change or other challenges.

"This provides a mechanism for the accumulation of variation in early development," Garfield says. The question now is whether particular events can cause that cryptic variation to rise to the surface, an idea that the Wray lab plans to test by exposing the urchins in the lab to environmental extremes.

"In the big picture, it's important to understand how organisms can go from a conserved developmental program to a sudden shift," Garfield says. "This says that there is genetic variation out there that could potentially be released without invoking any strange evolutionary events. It could mean some species aren't in as bad a shape as we think they are. There is worry that there is too little variation for them to adapt, but there may be cryptic variation out there just waiting to emerge." ▶



Sea urchin developmental gene networks show more variation than had been anticipated.

Interlocking Pieces

Signaling pathways are often thought of as linear events. But, of course, it doesn't really work that way. Single pathways carry out multiple activities. They branch, intersect and feed off of one another, and the pathways underlying cancer are no exception.

"The idea of a pathway as a long line of events leading up to some change in cell behavior is a convenience," said the IGSP's Erich Huang, Assistant Professor in the Department of Surgery. "We all understand they are not isolated circuits. Pathways talk to each other. It's more useful to think of them as modules that can interlock in different ways."

That means that changes in a single pathway will reverberate throughout the system. A drug designed to target one molecule will have what Huang refers to as "some fuzziness" in its effects. To really understand those effects and the biology of the cancer itself, you need to know how all those pieces fit together.

Researchers at the IGSP including Huang, Joe Nevins, Mike West, Joe Lucas and others have been working on methods to better characterize that biology and the true heterogeneity of the disease we call

cancer. A few years ago, they came up with a way to tie gene expression profiles back to the cancer pathways that were driving that tumor's progression.

More recently, they went a step further with a strategy that breaks up individual pathway signatures into biologically meaningful slices. A given pathway might be dissected into 20 different modules or sub-signatures, each with potentially important implications for the cancer.

"The method frees us from fixed linear models," Huang says. "It gives us more flexibility to look at useful variations."

Ultimately, Huang's goal is to create a library of those pathway modules representing all of the biological pathways with relevance to cancer. He likes to think of it as a huge box of Legos that, in different combinations and configurations, can represent the diversity and uniqueness of each individual cancer.

"We can start fitting those Legos together," he says. "Like one of those Star Wars kits, you can put them together to make the Millennium Falcon or the Star Destroyer. It's the same pieces, but they fit together to make different spaceships." ▶

"We all understand pathways are not isolated circuits. It's more useful to think of them as modules that can interlock in different ways."

—Erich Huang



Tag, You're It

Once upon a time, scientists thought they had the fundamental cell cycle pretty well figured out. A handful of genes called cyclins governed the carefully timed program of cell growth and division. But Associate Professor in Biology Steve Haase and colleagues threw a wrench in that understanding with a *Nature* study that surveyed the activity of thousands of genes at different time points across the cell cycle in mutant yeast lacking most of their cyclin genes.

Under the old models, the parade of gene activity should have come to an abrupt halt. Instead, while the yeast cells outwardly showed signs of the disruption and stopped dividing, nearly 70 percent of the periodic genes within them continued to turn on and off right on schedule.

Those results led Haase's team to a new model in which the cell cycle is controlled not by a few genes but by a network of influential genes that act as a kind of genetic tag team, one transcription factor activating the next in series. "The proposal challenged what the field had considered for a long time," Haase says.

His group has since confirmed that cyclic gene activity continues even when every last cyclin is lost, though the period of that activity does change somewhat. He is now working with quantitative heavyweights, including John Harer, Sayan

"It's a new way of thinking. It's not as simple as one or two or three genes interacting with each other."

—Steve Haase

Mukherjee, Alex Hartemink and David Schaeffer, to develop mathematical and statistical models to describe the network, a practice that is one of the hallmarks of a systems biology approach.

The end goal is a model thorough enough to predict how changes in a single gene in the network might alter the rest. The hope is that those models can ultimately point them back to experiments with the greatest likelihood of revealing important new details of a system that is fundamental to all life, from the yeast Haase works on all the way up to human cancer cells.

"As experimentalists, we've traditionally had to rely on intuition or do all possible experiments systematically," he says. "As the system gets bigger, it becomes more difficult to do that. The models can point us in the right direction." ▶

To Be, or Not to Be

For cells, life is a delicate balance. Each and every cell has a constant decision to make about its own fate – whether to grow, die, or just stay put. When cells lose control over themselves as they do in cancer, it can mean disaster.

Cells that go off the deep end and begin dividing uncontrollably have something in common: persistently high levels of a protein called MYC, says IGSP Investigator and Biomedical Engineer Lingchong You. MYC feeds into an important pathway that You's group found acts like an on-off switch for growth, with levels that are usually carefully adjusted based on external growth factors. In tumors, mutations can lead MYC to rise, short-circuiting the path from growth factors to this master switch.

Still, it hasn't been clear what happens when normal cells are made to over-produce MYC. To

“Noise isn't bad from the perspective of cells. It gives them the opportunity to make choices.”

—Lingchong You

get to the bottom of it, post-doctoral fellow Jeffrey Wong inserted extra MYC into otherwise normal cells with an engineered virus. But there was a problem. The cells took up the virus and the cancer-causing gene it carried in vastly different quantities – a detail largely ignored in previous studies. At first frustrated, You and Wong soon realized they had an opportunity. By quantifying MYC in each cell one-by-one and its individual response, they could calculate a dose-response curve.

The exercise led to a surprise. When MYC goes up, a master control gene for cell growth called E2f1 goes up too – for a while. But if MYC levels rise higher, E2f1 levels drop back down.

“The cells are very calculating,” You said. “Once MYC levels rise too high, they become suspicious.”

The finding exemplifies a growing focus in systems biology in the heterogeneity inherent in biological systems, a property You refers to as noise. “Biological noise is everywhere; If you look at different cells, they behave differently,” he says. There has been increasing interest in how that noise comes about, how it propagates through gene networks, and what its biological implications might be.

“Noise isn't bad from the perspective of cells,” You says. “It gives them the opportunity to make choices.”

The Right Combination

Jack Keene, James B. Duke Professor of Molecular Genetics and Microbiology, started sequencing virus genomes more than 35 years ago. Today, most viral pathogens have been fully sequenced, he says, but we still don't know much about how viruses make us sick. For that reason, he has always been skeptical about the many promises and expectations that swirled around the sequencing of the human genome.

“Even in the case of simple organisms with a few thousand nucleotides, we can't control them on that basis,” he says. “We can tinker with them, but the systematic understanding is still coming. You have to understand the genome, the proteins and where they come from, but it doesn't solve the problems we ultimately need to solve. That's a different kind of complexity; one that I think can be solved by simplifying principles.”

Keene turned his attention in the mid-1980s from viruses to the host cells they infect. His work in patients' autoimmune antibodies led to the discovery of RNA-binding proteins that are now known to be members of one of the largest families of proteins in the human genome. Those proteins bring order to the human genome by collecting RNA messages (mRNAs) that are to be translated into proteins that work together.

A decade ago, Keene proposed that those binding proteins provide an information management and distribution center for cells. These “RNA Regulons” help to explain how cells of every stripe can manage the varied tasks they must perform efficiently, by assembling RNA messages in different combinations.

“It's a combinatorial code of genetic information, that's really the point,” Keene says. “It's a way to make the 25,000 genes in the human genome go farther because there are many copies of each RNA message. It's economical while allowing new functions that evolve to be coordinated in kind.”

Under Keene's model, now evidenced to operate in all three domains of life, each of those organized collections of a few hundred RNA molecules becomes a small, modular subgenome unto itself. Ultimately, this layer of regulation and control over genes is critical for understanding how the networks that are much of the focus of systems biology function and how they change over time.

“Just like FedEx or UPS, cells have evolved ways to organize and coordinate information transfer after RNA leaves the genome,” Keene says. It's not enough to consider only which genes are actively copied from DNA into RNA.

“You have to understand the genome, the proteins and where they come from, but it doesn't solve the problems we ultimately need to solve. That's a different kind of complexity.”

—Jack Keene

A Place to Start

You might expect that genes would all have a precise beginning and end, but it turns out you'd be wrong. In fact, the enzymes that copy many genes into RNA will often start transcribing at any point over some window of the genome to produce messages that are either a bit longer or shorter. That pattern had been noted in mammals, and was recently confirmed for fruit flies in a *Nature Methods* report by the IGSP's Uwe Ohler and Jun Zhu.

Scientists had missed that before because they were relying only on the protein coding portions of genes. "The transcription start sites can be somewhat different from the part that codes for protein, so our knowledge had not been that comprehensive," Ohler explained.

Ohler and Zhu devised a new deep-sequencing method to find the places where genes truly begin. It turns out some genes do start in a more predictable fashion, while others, particularly housekeeping genes that cells express all the time, are transcribed in a less defined way. Those differences depend in part on particular regulatory sequences and on how tightly wound the DNA is around the proteins that package it into a condensed form known as chromatin, they've found.

The findings stem from Ohler's broader interest in understanding how genes are controlled. "How do you encode that each gene is expressed at the right moment and under the right conditions?" he asked.

As a computational biologist, Ohler approaches that overarching question at many levels, often relying on data generated by others in a variety of species to get at the problem from different angles. With international collaborators at the Max Planck Institute of Molecular Cell Biology and Genetics in Germany, he is tackling the problem on an evolutionary scale.

The research team began by characterizing gene activity in fruit fly embryos representing different species as they developed over time. Those profiles showed that the expression of some



A mosaic of a fruit fly made up of images of fruit fly embryos at different developmental stages. Credit: Pavel Tomancak

genes doesn't change much from one species to the next while for others it shows considerable divergence. "We wanted to understand how that comes to be," by tracing those differences back to the regulatory regions that control them, Ohler said.

That effort turned up an interesting pattern just reported in *Nature* that bears on a much-debated observation dating back to the 19th century. In those days, scientists realized that embryos representing completely different species – mouse and man, say – can look remarkably similar at certain points in development. That pattern of resemblance tends to follow an hourglass model. Developing embryos go from looking different, to looking the same, to looking as different as a human and a mouse.

It turns out that pattern holds up at the genomic level, the new evidence shows. Genes expressed early and late in development tend to vary more in their activity than those expressed somewhere in between. "Using modern gene expression data and comparative genomics, we've suddenly answered a developmental biology question that has been around for 150 years," Ohler said.

Ohler is now working on methods to analyze the associated regulatory bits in the genome and how they vary from one species to the next. "People know how to compare coding sequences but, for the regulatory code, it's much more murky," he says.

In keeping with his Duke colleagues, Ohler's ultimate goal is a global view of whole biological systems, complete with all their moving parts. "We want to understand not one gene in isolation, but how everything in the genome comes together to make a complex, developing organism." ▶

“ Using modern gene expression data and comparative genomics, we've suddenly answered a developmental biology question that has been around for 150 years. ”

—Uwe Ohler

Mining the Mustard

A weedy relative of cabbage and mustard, the small flowering plant known as *Arabidopsis thaliana* doesn't look all that impressive. But it turns out we have a lot to learn from them. *Arabidopsis* has been a laboratory mainstay for decades in part because of its small genome and rapid growth. It was also the first plant to have its genome fully sequenced. Researchers have since been marching through the mustard genome in an effort to catalog the function of each and every gene and the plant now boasts one of the best quality genome sequences around.

In Philip Benfey's estimation, *Arabidopsis* is clearly the best species to advance plant systems biology. And more often than not, the biological insights made in the model plant are likely to have relevance to other organisms.

"Just about everything in biology that once seemed particular sooner or later proves to be more general," Benfey has said. And that makes sense. Once fundamental biological processes got worked out over the course of evolutionary time, organisms have tended to keep those mechanisms around.

But if there can still be any doubt about the promise of exploring *Arabidopsis* through the lens of genomics and systems biology, one need look no further than the work reported out of the Benfey lab just in the last year.

His team focuses on the intricate details of *Arabidopsis* root growth and development. Those careful studies, in which they typically sort out individual cell types and consider each in turn, have led the researchers to all kinds of fundamental discoveries.

In a report in *Nature* last April, Benfey's team found something rather unexpected regarding tiny bits of genetic material known as microRNA. Those microRNAs have drawn interest since their discovery in the early 1990s as important regulators of gene activity within cells.

But the *Arabidopsis* work showed that microRNAs aren't limited to controlling the activity of genes within a given cell; they can also move from one cell to another to send signals that

influence gene expression on a broader scale. It's still not clear how the microRNAs travel, but it appears in the case of *Arabidopsis* that their mobility allows them to play an important developmental role in sharpening the boundaries that define one plant tissue from another.

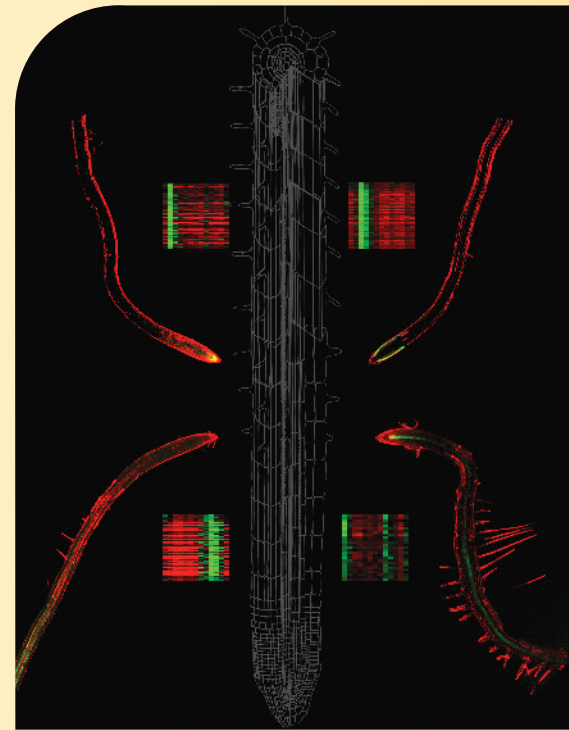
"To our knowledge, this is the first solid evidence that microRNAs can move from one cell to another," Benfey said. The finding added microRNA to the list of mobile molecules, including hormones, proteins and other forms of small RNA, that allow for essential communication between cells in the process of organ development.

Soon after, another report revealed a surprisingly simple connection between growth and development — a critical link ensuring plants and animals end up with just the right number of cells and in all the right places. It turned out that a single developmental protein that had been well-studied by Benfey and his colleagues controls the activity of other well-known genes involved in cell division.

In another surprise, Benfey's group uncovered an unexpected parallel between the development of an animal's spinal column and a plant's root system: Both are controlled by a "molecular clock" that governs a regular spatial pattern of development. In *Arabidopsis*, that wave of gene expression travels up from the tip of the roots and reaches its peak every six hours. That wave includes thousands of genes, including two separate groups that trade off their activity; when one group is high, the other is low and vice versa.

In vertebrates, the mechanism ensures that each species ends up with the correct size and number of vertebrae. In the case of the plant roots, those peaks mark the spots where root branches can later form.

"It appears there are similar underlying processes in both plants and animals, suggesting that the number of ways the problem can be solved is limited," Benfey said. "This must be a very efficient solution."



Detailed genomic studies of individual cells in the *Arabidopsis* root have led to many fundamental biological insights.

Finally, a genome-wide search for genes that come 'on' precisely when root cells transition from proliferation to differentiation and then turn 'off' again just as quickly turned up a single gene that can shift the balance. When that one gene is mutated, it produces plants that just keep growing.

That discovery may have real practical application for perennial grasses grown for biofuels, Benfey says. Those grasses usually can't be harvested until their second or third year, and a method to improve root growth could really speed up biofuels production.

Ultimately, though, it all comes back to the biology. "These approaches yield biological insights," Benfey says. "That's the common theme." ▶

“Just about everything in biology that once seemed particular sooner or later proves to be more general.”

—Philip Benfey

Take a Closer Look

Despite all the years of laboratory study, the rather diminutive yeast still holds many lessons if you look at them closely enough. Not long ago, Assistant Professor of Biology Paul Magwene's lab sequenced 16 wild yeast strains isolated from various natural settings – woodlands, vineyards, and patients in the clinic – to find that the many of them were far from being devoid of genetic variation as had been anticipated.

It now appears that variation can be traced to a particular strain's lifestyle. Those that live in close association with us humans tend to harbor more genetic diversity. Human activities apparently bring different strains together, giving them the opportunity to have sex. But most of the time those strains don't have sex at all. They simply produce clones of themselves, a habit that allows them to keep whatever diversity they had to start. When yeast strains are left to their own devices, they do have sex, often with individuals just like themselves. Any genetic diversity they might have had is rapidly lost.

At the level of molecules, those different strategies are tied to the activity of a core growth pathway known as the cAMP pathway and, through a combination of experimentation and mathematical modeling, Magwene and collaborators Kevin Gonzales and David Schaeffer in the Math Department, are learning new and unexpected things about that too.

“Everybody talks about interdisciplinarity, but it's not easy to do.”

—Paul Magwene

“cAMP is a conserved signaling mechanism and we know all the players,” Magwene said. “Surprisingly from my perspective, we didn't seem to know a lot about the dynamics.”

Gonzales and Schaeffer are distilling what's known about the pathway into a series of differential equations capable of reproducing behaviors of normal and mutant yeast in the lab. Those equations suggested something that had never been reported before: under some conditions cAMP levels should waver up and down, producing what they call oscillatory dynamics.

Lo and behold, when they looked at cAMP dynamics in different strains and over a longer time period, those predicted dynamics showed up. Without the close collaboration with those working outside of his field – a hallmark of a systems biology approach – Magwene says none of it would have been possible.

“Everybody talks about interdisciplinarity, but it's not easy to do,” Magwene said. “You need the right context – having access to and regular interactions with people from different disciplines who are immersed in thinking about similar sorts of problems.” ▶

Battle of the Sexes

In the developing embryo, the gonads have a very important decision to make: whether to become testes or ovaries. That makes them rather special.

“The gonad is the only organ that forms and then differentiates as two completely different things,” says Blanche Capel, a professor in the Department of Cell Biology. “The fact that it determines our sex also captures the imagination.”

Different organisms determine sex in different ways. In mammals, including us, the key factor is genetics. Those who inherit a Y chromosome from their father will become males and those who inherit an X will become females. In turtles, temperature reigns. In some fish, one dominant male can prevent its fellow fish from following the male path. Some species even experience partial sex reversals over the course of a year as the seasons change.

Despite those differences in the master switches that send an individual down one path or

another, Capel suspects that a similar underlying network of genes might be at play. And even in the seemingly simple XY system, ultimately controlled by a single gene on the Y chromosome, that network involves what Capel has referred to as “complex antagonistic signaling pathways that stage the battle of the sexes in the bipotential gonad.”

Capel's lab got its start in sex determination with more classic developmental biology experiments, work that is still ongoing today. She says it has really only been in the last couple of years that they've begun to branch out, incorporating genomic technologies and a systems biology approach in mice. To the extent possible, they aim to do similar work in red-eared slider turtles as well, with plans to sequence the transcriptomes of embryonic turtles in the IGSP's Genome Sequencing and Analysis Facility both before and after sex is determined.

Already, the mouse work has led to new insight.

For instance, before the gonad has made its decision one way or another, it appears that there are two, partially overlapping networks running simultaneously, one male and the other female.

“It's exciting to see the science evolve,” Capel said. “We're beginning to get a picture of the gonad at a very different level than people have ever had before while looking at it one gene at a time. We're attempting to understand it as a system.” ▶

“We're beginning to get a picture of the gonad at a very different level than people have ever had before.”

—Blanche Capel



The Institute for Genome Sciences & Policy

Duke University
CIEMAS
101 Science Drive
Box 3382
Durham, NC 27708
www.genome.duke.edu



Systems Biology *S e m i n a r*

Systems biology at Duke was nurtured through regular open meetings interleaving research presentations with tutorials and journal clubs. The basic character of those meetings remains the same today. All are invited, and so-called dumb questions are not only accepted but welcomed. These questions are critical to educating ourselves and forming crucial links between disciplines.

The seminars are held on Wednesdays at 2:30 in French 4233. Details at genome.duke.edu/centers/csb/seminars/

Mark your calendars for the inaugural **Duke Synthetic Biology Symposium** on April 16, 2011. The event is co-sponsored by the IGSP and its Center for Systems Biology, the Center for Biomolecular and Tissue Engineering, the Pratt School of Engineering and the Department of Biomedical Engineering. Details at biology.duke.edu/synthetictbiology/

We've got an event for you!

In addition to the Systems Biology Seminar, the IGSP hosts or sponsors a wide selection of regular lectures, meetings and workshops throughout the academic year:

Computational Biology Seminar, Mondays @ 11:00

Cancer Genomics Meeting, Mondays @ 1:00

University Program in Genetics & Genomics Tuesday Seminar Series, Tuesdays @ 12:30

Genome Biology Meeting, Tuesdays @ 4:15

Genomic and Personalized Medicine Forum, Thursdays @ 9:00

Synthetic Biology Group Meeting, the 2nd Friday of the month @ 3:00

Science & Society Journal Club, the 3rd Friday of the month @ 11:30

Plus, educational workshops, including **Genome Academy, Career Series Dinners** and more. Visit genome.duke.edu for details and the latest updates.