

MORE THAN FATE:

Computation Addresses Hot Topics in Stem Cell Research

By Katharine Miller

To the casual observer, stem cells offer the almost magical promise of—Voila!—turning into exactly the kind of cell needed to repair an injured spinal cord or replace a damaged organ. And despite the political issues that swirl around the topic, new research findings fuel the public's hope that the stem cell miracle is right around the corner. Since 2005, scientists have gotten better and better at converting adult skin cells into pluripotent stem cells capable of becoming any cell-type in the body.

But beneath these exciting results lies a far more subtle truth: Although researchers can produce desired cell lines in the lab, they don't always understand the underlying mechanisms. "It works, but we don't necessarily know

why it works," says **Ingo Roeder, PhD**, group leader at the Institute for Medical Informatics, Statistics and Epidemiology at the University of Leipzig in Germany.

Stem cells are complex creatures, responding to external and internal cues with an array of cellular changes, including alterations in gene expression, DNA methylation, alternative splicing, microRNA (miRNA) expression, and post-translational modification of proteins. Researchers need to understand these intricacies before they can control stem cells for clinical purposes. But as researchers start to track all of these changes in cells over time, the vast quantity of accumulating data overwhelms the human mind. Researchers can even lose track of what hypotheses they are testing. The only way to make sense of it all is with computation.

"Computation allows you to distin-

Above: Differentiated cells (stained red) have developed to surround the tightly packed colony of smaller pluripotent cells (stained green). Photo courtesy of John Butler, Dr. Jeanne Lawrence, Dept of Cell Biology, UMASS Medical School.

guish between hypotheses in systems where we don't always have all the information we need," says **Peter Zandstra, PhD**, professor of biomaterials and bioengineering at the University of Toronto in Canada.

The scenario runs something like this: As experimental results accumulate, stem cell researchers start developing theories about why stem cells do what they do. Computational biologists then develop computer or mathematical models to examine the theories in a rigorous way, to guide further experiments. "That's in my mind where a lot

of the power is,” Zandstra says.

Using computational models, researchers are gaining traction toward understanding what makes a stem cell a stem cell; how gene expression

leagues built an enormous database (which they call “the stem cell matrix”) that contains data on gene expression, microRNA expression, DNA sequencing, and epigenetics

rocks,” Loring says.

Since the work was published in *Nature* in 2008, the clusters have held up using additional data. Loring’s team has now applied their analysis to more

“Computation allows you to distinguish between hypotheses in systems where we don’t always have all the information we need,” says Peter Zandstra.

drives stem cell differentiation; why studying stem cell heterogeneity is important; and, ultimately, how stem cells control their fate.

WHO AM I: The Stem Cell Identity Crisis

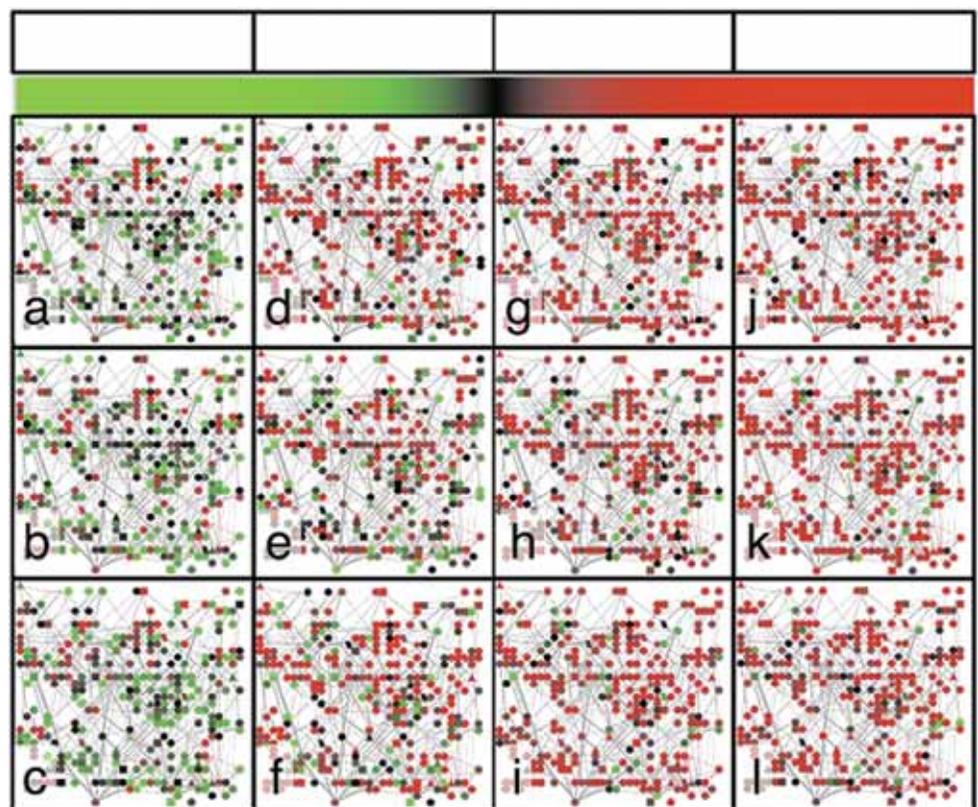
Stem cells seem to know who they are, but it can be hard for humans to tell them apart. Stem cells come in two general types: pluripotent stem cells that can give rise to all cell types in the body, and multipotent stem cells (such as those in bone marrow or the brain) that have a smaller repertoire of options. There are also adult cells that have been induced to resemble stem cells; and cancerous cells that exhibit stem-like traits. And there are many shades of gray in between—stem cells in the process of differentiating; or adult cells in the process of reverting into stem cells. So when stem cell researchers work with cell cultures, it’s not always an easy matter to determine what kind of cells lie before them. They need a way to determine whether they’ve successfully pushed stem cells to differentiate or induced differentiated cells to become pluri- or multipotent. In essence, they need a test for stem cell status.

Jeanne Loring, PhD, professor of developmental neurobiology and director of the Center for Regenerative Medicine at The Scripps Research Institute in La Jolla, California, is taking a bioinformatics approach to address this problem. She and her col-

leagues built an enormous database (which they call “the stem cell matrix”) that contains data on gene expression, microRNA expression, DNA sequencing, and epigenetics (DNA methylation), among other things. Using the data from 22 samples and a machine-learning algorithm, they taught a computer how to identify stem cells. When applied to 66 test samples, the algorithm clearly separated pluripotent stem cells into a class by themselves. “They are a different category from all other cells—as distinguishable as white rocks from black

than 500 samples—with data on expression of 40,000 genes per sample and 37,000 DNA methylation sites per sample. “The more information we get, the simpler the answer is,” Loring says. “We get rid of the noise.”

Loring’s lab plans to offer a simple method to help scientists determine what cell type they are looking at right now. “Researchers will send us a sam-



Using a program called Matisse, developed by researchers in Israel, Loring and her colleagues superimposed gene expression data from her lab’s stem cell matrix on known protein-protein interactions. This filtered the data down to the key set of interacting proteins that were commonly all up-regulated in pluripotent stem cells—a network they dubbed the PluriNet. This figure displays the activity of the PluriNet in three different samples of (in columns from left to right) four different cell types: nonpluripotent stem cells, tumour-derived (cancerous) pluripotent cells, induced pluripotent stem cells and pluripotent embryonic stem cells. The pluripotent cell lines (right two columns) resemble one another and are quite different from the other two). Reprinted with permission from McMillan Publishers, Ltd, Muller, FJ, et al., *Regulatory networks define phenotypic classes of human stem cell lines*, *Nature* 455, 401-405 (18 September 2008).

ple or gene expression data, and we can tell them which category their cells are most like,” she says.

THE ATTRACTOR LANDSCAPE: The Lay of the Land for Stem Cells

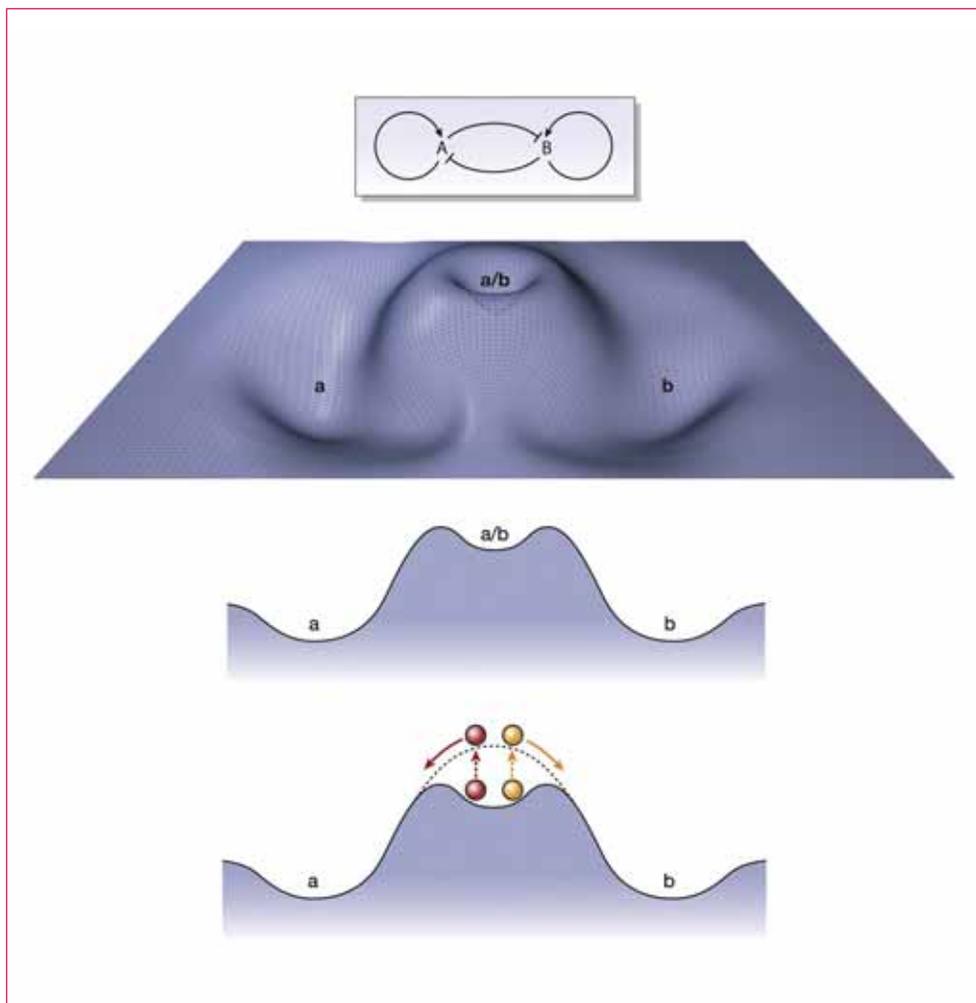
For more than 40 years, cell biologists have described stem cell differentiation in terms of a metaphorical energy landscape. The cell’s gene expression state—essentially the transcriptome—can be stable in multiple different combinations. When it finds a stable state, it stays there, like a marble stuck in a valley on the landscape. The starting “well” is the pluripotent stem cell. Certain driving forces then push the cell up and out of those wells and across ridges into new low areas. Called attractors, these low areas represent specific differentiated states, such as neurons or blood cells.

Many computational researchers still consider this depiction useful only as a metaphor. But a few are taking it further. “It’s actually a very mathematical thing,” says **Sui Huang, PhD**, associate professor of biology at the University of Calgary in Alberta, Canada. Within every cell, there’s a network of genes interacting with other genes. These interactions generate gene expression patterns that define the cell’s state—i.e., whether it’s a stem cell or has differentiated into some other kind of cell.

Some possible states are more likely to be stable than others. For example, if gene A inhibits gene B, then a pattern where A and B are both highly expressed is very unlikely to be stable. So, in theory, if you understood the wiring diagram for all gene interactions, including which genes inhibit or activate which other genes, you could predict the likelihood that the network permits a particular gene expression pattern. And, says Huang, “that probability would give you the derivation of the landscape.” There is a caveat, Huang says. “From a physics point of view, it’s not really an energy landscape because living systems are not equilibrium systems, but the intuition of landscapes is wonderful.”

Of course, to compute the landscape, Huang says, you would have to know the details of the wiring diagram—which genes activate each other

The cell’s gene expression state—essentially the transcriptome—can be stable in multiple different combinations. When it finds a stable state, it stays there, like a marble stuck in a valley on the landscape.



In this gene regulatory network, two transcription factors (A and B) activate their own expression and mutually inhibit each other. A mathematical model of this network produces a landscape with three “attractors”—states of gene expression that are stable. One (a/b) in which there is low expression of both A and B, would be a stem cell-like state, whereas the valleys a or b constitute stable states in which A or B are exclusively expressed. If the stem cell’s state is somehow destabilized, the upper valley becomes a hilltop (dotted black line) rather than a valley, inducing the cell’s state to move toward one of the differentiated (committed) states. If A and B expression fluctuate within the stem cell valley, then those near the lip of the valley will be the first to commit because they are already prone to flow in a particular direction. Reprinted from Cell Stem Cell, Vol 4 issue 5, Enver, T., et al., Stem Cell States, Fates, and the Rules of Attraction, pages 387-397 (2009), with permission from Elsevier.

and how. And, he says, “getting that wiring diagram is a big, big step.”

So far, Huang and his colleagues have modeled the landscape with a two-gene interaction network. The genes function as a stem cell switch for hematopoietic (blood-forming) stem cells: transcription factor PU.1 drives them to become white blood cells, while GATA1 drives them to become red blood cells. Each transcription factor activates itself and inhibits the other.

For a network with two genes, the third dimension is the elevation, which equals the probability of each possible expression pattern of PU.1 and

GATA1. This elevation is very hard to compute even with just two genes, Huang says. “You cannot derive a mathematical equation to give the shape of the landscape,” he says. “You have to use brute force.”

Looking at more than two genes would require an even greater number of dimensions. “If you have 100 genes then your state space has 100 dimensions and the elevation is the 101st dimension,” Huang says. “That’s hard to picture, and computing that landscape would be very intense.”

But even Huang’s two-component model provides useful insights. For example, his team modeled the trajec-

tories that the cell takes from a particular undifferentiated state to the differentiated state. Surprisingly, it’s not necessarily a straight line; the levels of PU.1 and GATA1 fluctuate and even loop around before the cell moves toward the attractor. By modeling this, and comparing the *in silico* trajectories to those determined experimentally, the researchers better understand how the two genes are interacting in the cell to decide its fate.

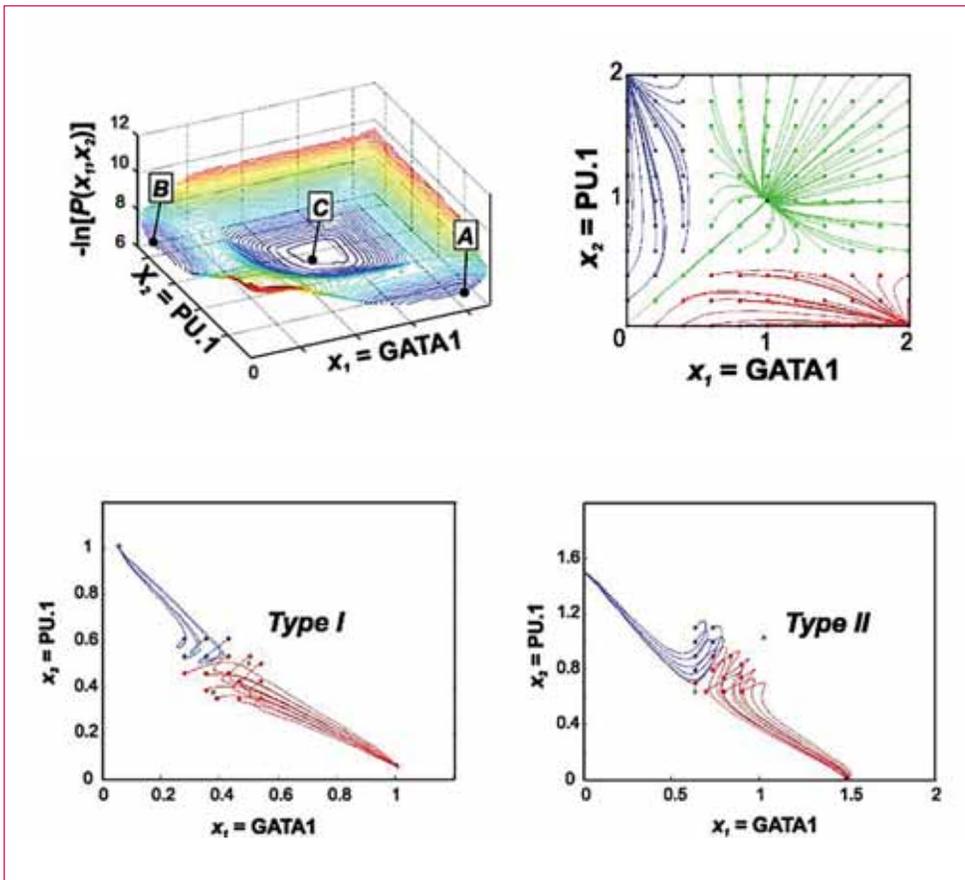
INDIVIDUALS MATTER: Understanding Stem Cell Heterogeneity

Much of stem cell biology relies on population statistics—for example, the average gene expression levels of a vast

“Outliers matter in biology,” says Sui Huang. “All you need is one cell behaving differently and it has consequences for the organism.”

numbers of cells grown in a culture. But stem cells exhibit a surprising degree of individuality even within such cultures. Gene expression varies greatly; cells divide asymmetrically; some cells die out while others reproduce themselves endlessly. Moreover, adding certain chemical cues to stem cells will induce some, but not others, to differentiate.

“Outliers matter in biology,” Huang says. “Science has a tendency to operate with averages—average populations, average females, average males, but individuals are important. All you need is one cell behaving differently and it has consequences for the organism.”



*In this illustration of the attractor landscape for a two-gene network, (top left) the blue wells representing hematopoietic stem cell state C as well as the attractor states A and B for erythroid and myeloid (red and white blood cell) cells respectively. At the top right, the state space for the PU.1/GATA1 system shows how the cell state is driven by the network dynamics from various possible starting states (points on the grid) toward the attractors: Blue lines move toward the myeloid state, red toward the erythroid state, and green toward the metastable hematopoietic stem cell state. Adjustments to the model parameters will shift the size of the area represented by each color. According to one theory, differentiation occurs when the multipotent state (green) becomes unstable—essentially disappearing so that the cell must choose between the two possible fates. As shown at bottom, Huang and his colleagues also modeled the trajectories that the cell might take from an initial stem cell state (in the center of the state space) to each corner, given different sets of initial conditions. These trajectories were then compared to actual trajectories observed in the lab. Reprinted from *Developmental Biology*, 305:695-713 (2007), Huang, S., et al., *Bifurcation dynamics in lineage-commitment in bipotent progenitor cells*, with permission from Elsevier.*

Heterogeneity also means that any given attractor is actually a cloud, rather than a point, on the landscape, Huang says. “So we need to have the statistics for thousands of individual cells to get the landscape,” he says. “We need to follow individual cells.”

To do that, researchers turn to microfluidic devices that can rigorously control the environment surrounding individual cells without washing them away or moving them around. In time-lapse experiments, a camera can capture gene expression levels in individual cells while also tracking cells as they divide, die, differentiate, or remain pluripotent. The data retrieved can help confirm or deny predictions from computational models.

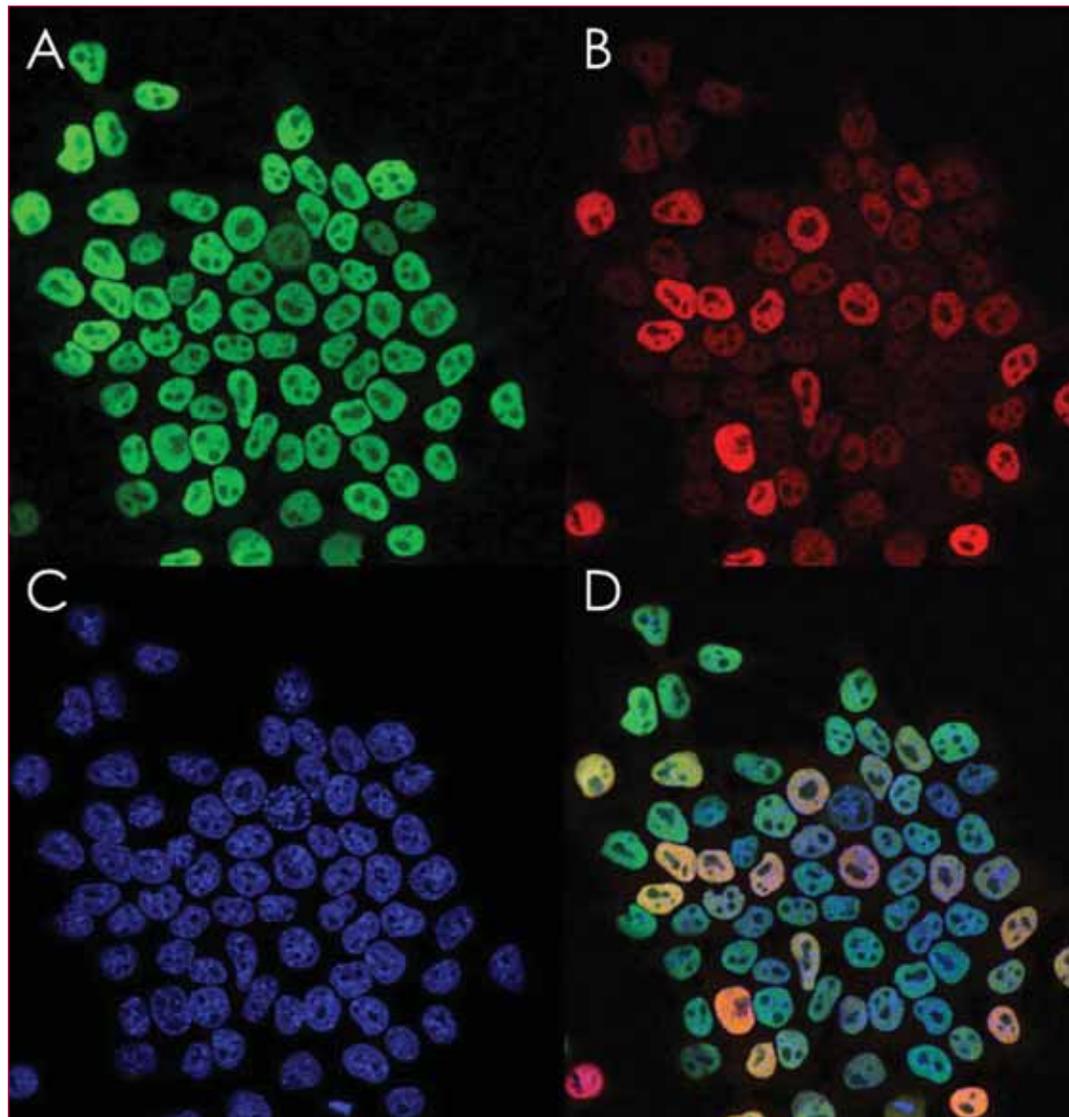
For example, Ingo Roeder used a system of differential equations to model the wide fluctuations observed in *Nanog* expression in individual mouse embryonic stem (ES) cells. The work (unpublished) suggests two possible explanations: either cells themselves fluctuate between two possible stable states (induced by random perturbations—essentially noise); or the state itself is oscillating

(such that the state is never really stable). There is no theoretical way to distinguish between these two scenarios based on average population statistics, Roeder says. Rather, experimentalists will need to monitor temporal changes in *Nanog* in individual cells over time.

Monitoring individual cells over time can also generate the cells’ genealogical trees. But, says Roeder, “There is currently no set way to computationally analyze those genealogical

his colleagues simulated an array of possible cell genealogical trees *in silico*. The cells can self-renew, die, or differentiate. To be realistic, the simulations

ical tree. The computer can distinguish proliferating cells from cells in a steady state or in decline, and can recognize asymmetry. Going forward, the goal is



“One of the major advantages of computer modeling is that you can try lots of different scenarios and then narrow down the possibilities for explaining certain behavior,” Ingo Roeder says.

trees.” So Roeder decided to get a jump on that problem before collecting data. Using a computer model of hematopoietic stem cell organization, Roeder and

include random noise.

The simulations showed that changes in the growth conditions (*in silico*) altered the shape of the genealogical

to determine whether different cell scenarios generate unique tree “signatures” that a computer can spot. These will have to be validated against the real

genealogies of cells. “With the experimental data, we will see the trees and use the simulations to estimate back to learn the underlying mechanism,” Roeder says.

University of California, Berkeley. So modelers build networks from what they know about interactions in the stem cell and then set the models in motion to see what happens.

with information about networks in other systems (such as in yeast, which is well understood), hypothesize the interactions that might be occurring, and then mesh that with all kinds of data being collected from stem cell systems, including protein expression data, miRNA expression levels, protein post-translational modifications, protein phosphorylation, signal transduction, and, increasingly, any or all of these types of data as a function of time.

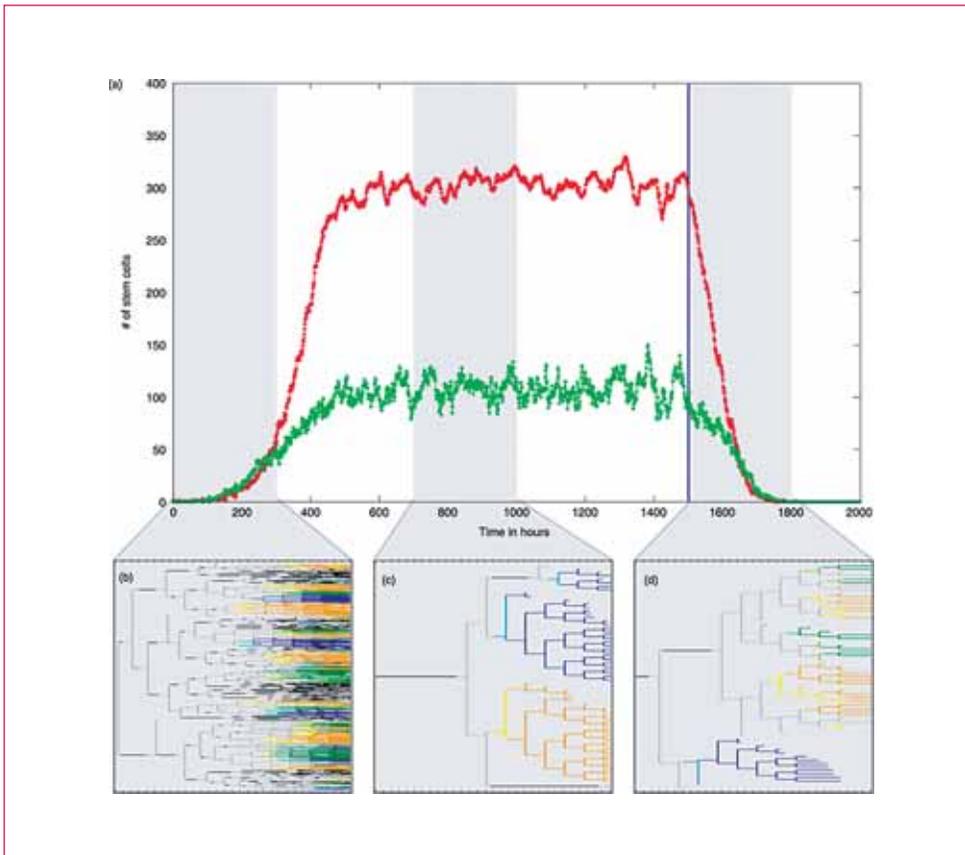
In a 2004 paper, Schaffer and his colleagues created a model to explore the dynamic behavior of the sonic hedgehog (Shh) gene regulatory network—a network known to function as a cell fate switch in certain contexts. The model used differential equations to track the rates of change in concentrations of network participants as well as the rates of protein synthesis and degradation. The model showed that the system functioned as a digital all-or-nothing switch that is not easily reversed.

In 2009, Schaffer’s team also modeled the Notch signaling pathway, known for its involvement in cell fate decisions during development and adulthood. They found that the Notch system also acts as a bistable switch, but they identified a factor that could change the system into an oscillator. Thus, the Notch system can be adjusted to exhibit different behaviors depending on the context. The work is currently being validated experimentally.

But these models do not necessarily represent the absolute truth. As so often happens in science, new information can come to light, requiring changes in the model. In a 2006 paper, **Carsten Peterson, PhD**, professor of biological physics at Lund University in Sweden, and his colleagues modeled three key transcription factors involved in embryonic stem (ES) cell self-renewal—Nanog, Sox2, and Oct4. The model showed that the three could—on their own—function as a bistable switch to maintain stem cell pluripotency.

Now, it appears that Oct4 activation is just an early step in the process, triggering the opening up of the chromatin region around Nanog and several newly identified transcription factors that play a key role in the switch. “On the very top you have these epigenetic things happening,” Peterson says. “That adds another dimension to the whole modeling perspective.”

Besides extending the model to



This chart shows the population dynamics of proliferating (red) and quiescent (green) cells as they move through three different scenarios: expansion, homeostasis, and terminal differentiation (i.e., dying off over time). Characteristic genealogies for each scenario are shown below. To the human eye, these genealogies are clearly different. Roeder and his colleagues sought to train a computer to spot distinct tree signatures. Reprinted with permission from John Wiley & Co., Glauche, I., et al., A novel view on stem cell development: analysing the shape of cellular genealogies, Cell Proliferation 42, 248-263 (2009).

For stem cell research, the benefits of this work may still be a ways off. But ultimately, Roeder says, “One of the major advantages of computer modeling is that you can try lots of different scenarios and then narrow down the possibilities for explaining certain behavior.”

IT’S FATE: Switches and Beyond

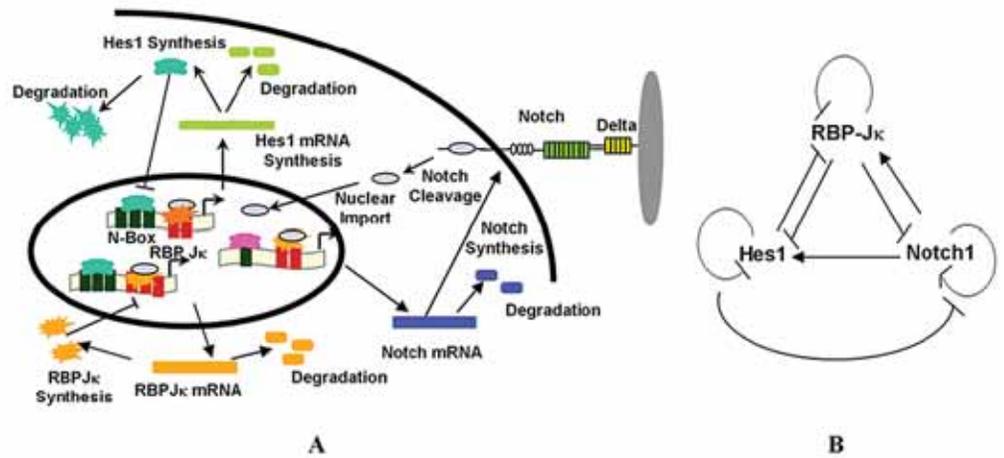
Many computer models of stem cells focus on the question of fate: How does the stem cell decide whether to remain pluripotent or differentiate into another kind of cell? “The most useful models recognize that decisions inside the stem cell are collective decisions of networks of interacting biological molecules,” says **David Schaffer, PhD**, professor of chemical engineering, bioengineering, and neuroscience at the

“A model is really a statement of hypothesis that aggregates our knowledge of how the system behaves,” Schaffer says. “You’re either right or not right. And when you compare the model predictions to experimental data, if you’re not right, then you know you’re missing something. That then motivates experiments to determine what you don’t understand.”

It’s often an iterative process, Schaffer says. “The model summarizes what we know but also guides experimentation so we can best learn more about the system experimentally.”

Until fairly recently, it was difficult to construct models of stem cells because there weren’t enough data available. But that is changing, Schaffer says. Now, to build a model of a network inside a stem cell, one can start

In this schematic of the Notch1-RBP-Jk-Hes1 signaling network (left), each arrow represents a term or event in the differential equation model including transcription, translation, mRNA and protein degradation, nuclear import, TF binding, receptor-ligand binding and receptor processing. At right, a schematic of the positive and negative feedback loops of the Notch1-RBP-Jk-Hes1 network. (-) represents repression and (-) represents activation of target genes. Reprinted from Agrawal, S., et al., *Computational Models of the Notch Network Elucidate Mechanisms of Context-dependent Signaling*, PLoS Computational Biology, May 2009.



include epigenetics, Peterson says, they are also trying to include mechanical interactions that play a role in ES cell differentiation and migration. “It turns out that mechanics are not negligible,” Peterson says.

When ES cells are inside the egg, some start to change into endoderm—the cells that form an outer shell around the inner ES cells. But initially, the cells that are changing are “like salt and pepper,” Peterson says. “They are all over the place.” To understand how the endoderm develops, he added different adhesion properties to the two types of cells in his model. And the computational result matched the experimentally observations: the endoderm cells move out, leaving the stem cells inside. Friction alone, without any influence from chemical cues, was enough to properly separate the two different types of cells. The next step will be to determine whether ES and endoderm cells actually exhibit different adhesion qualities and what genes cause those traits to develop.

Peterson and his colleagues are also modeling aspects of the hematopoietic stem cell system and finding interesting features that resemble locks. After the straightforward PU.1/GATA1 switch has been activated (described above), the hematopoietic system gets more complex. “Here’s where a mathematical model can help,” says Peterson. His model, published in 2009 in *PLoS Computational Biology*, suggests that downstream genetic players interact with one another and also send feedback to the PU.1/GATA1 switch, preventing changes in previ-

ously made decisions. “There have to be locks on the way down to make sure it’s irreversible,” he says. And it’s crucial to understand that irreversibility if researchers want to induce hematopoietic stem cells from differentiated blood cells.

NEIGHBORS MATTER: Modeling Stem Cell Interactions

Many models of stem cell switches look at the circuitry inside the switch without considering what threw the switch in the first place. But stem cells’ fate decisions depend, at least in part, on changes in the environment. Zandstra and his colleagues are modeling one key environmental component—cell-cell interactions—within the hematopoietic system.

The hematopoietic system is quite remarkable. Every day, hematopoietic stem cells in the bone marrow produce tens of millions of red blood cells as well as an appropriate number of white blood cells and platelets—all the different cellular components of blood. To be so reliable day in and day out, researchers believe the system must—at least in part—be tightly controlled by soluble factors secreted by blood cells. Because the process is poorly understood, researchers have a hard time growing hematopoietic cells in culture—a prerequisite to further research.

In a 2009 paper, Zandstra’s team made an initial foray into modeling how cell-to-cell interactions control hematopoietic self-renewal and differentiation. “We’ve built theoretical models of feedback systems where stem cells give rise to progeny through a series of fate decisions,” he says. At this point, he says, “We’re starting to understand the structure and connectivity of the cell-to-cell networks and what determines whether the stem cell population proliferates or differentiates.”

In addition, Zandstra’s team developed a way to test the model in a cell culture system by removing different

“We’re starting to understand the structure and connectivity of the cell-to-cell networks and what determines whether the stem cell population proliferates or differentiates,” Zandstra says.

cell types along the way—cutting out various feedback loops. “This has been very fruitful,” he says. “By understanding the intercellular networks and controlling them, we can

grow these cells far better than you could before.”

The underlying principles in Zandstra’s model should be applicable to stem cell systems beyond blood.

PREPARING FOR THE CLINIC

Although computational modelling of stem cells might not directly lead to therapies that treat Parkinson’s disease or Alzheimer’s, Zandstra says, the models help weed through potential solutions to find

“Once you have good models of how cells are maintained as well as transition or differentiate, you can start to think about how the various parts of the network are druggable,” Schaffer says.

those with the greatest impact. By understanding how stem cells make decisions, we gain the ability to control those decisions. “You can’t get the clinical outcomes without the increased control,” he says.

Schaffer agrees. “I really view my job as measure, model, manipulate,” he says. “Once you have good models of how cells are maintained as well as transition or differentiate, you can start to think about how the various parts of the network are druggable.” □

STEM CELL DIVERSITY AND DRUG TESTING

Someday soon, pharmaceutical companies will be converting stem cells into liver cells they can use for testing drug toxicity, says Loring. “It’s the wave of the future. There’s no better way to test a drug on liver than to grow liver cells in a dish and dump the drug on them.” This approach could help drug companies better understand variations in the way people react to drugs. But there’s a problem looming, Loring says: Almost all of the preclinical work with embryonic stem (ES) cells uses Caucasian cell lines.

In a 2009 paper published in *Nature Methods*, Loring and her colleagues used a Bayesian analysis of ES cell genotypes to determine the ethnic background of existing ES cell lines. What they found—the dominance of Caucasian cell lines—springs from the cells’ source in in vitro fertilization (IVF) clinics. “Embryonic stem cells made from embryos discarded at IVF clinics are almost all Caucasian and East Asian,” Loring says. There are almost no African stem cells, she says. “So our soapbox is that if the pharmaceutical industry is going to start using pluripotent stem cells, it needs to incorporate diversity.”

“This is more important than anything I’ve ever done,” Loring says. If all the preclinical work is done on Caucasian cell lines, then the pharmaceutical companies might release drugs that are toxic to some people and don’t work on others. Loring hopes her paper puts some pressure on pharmaceutical companies. “They need early assays for toxicity that capture the diversity of people.” The *Nature Methods* paper took a first step in that direction, publishing the creation of an induced stem cell line from skin cells of a Yoruba (Nigerian) individual.



This map shows the geographic location of the source institution (geographic origin) and the ethnic origin for 52 pluripotent embryonic stem cell lines. Reprinted with permission from *McMillan Publishers, Ltd., Laurent, LC, et al., Restricted ethnic diversity in human pluripotent stem cell lines (both embryonic and induced), Nature Methods 7, 6-7 (2010).*