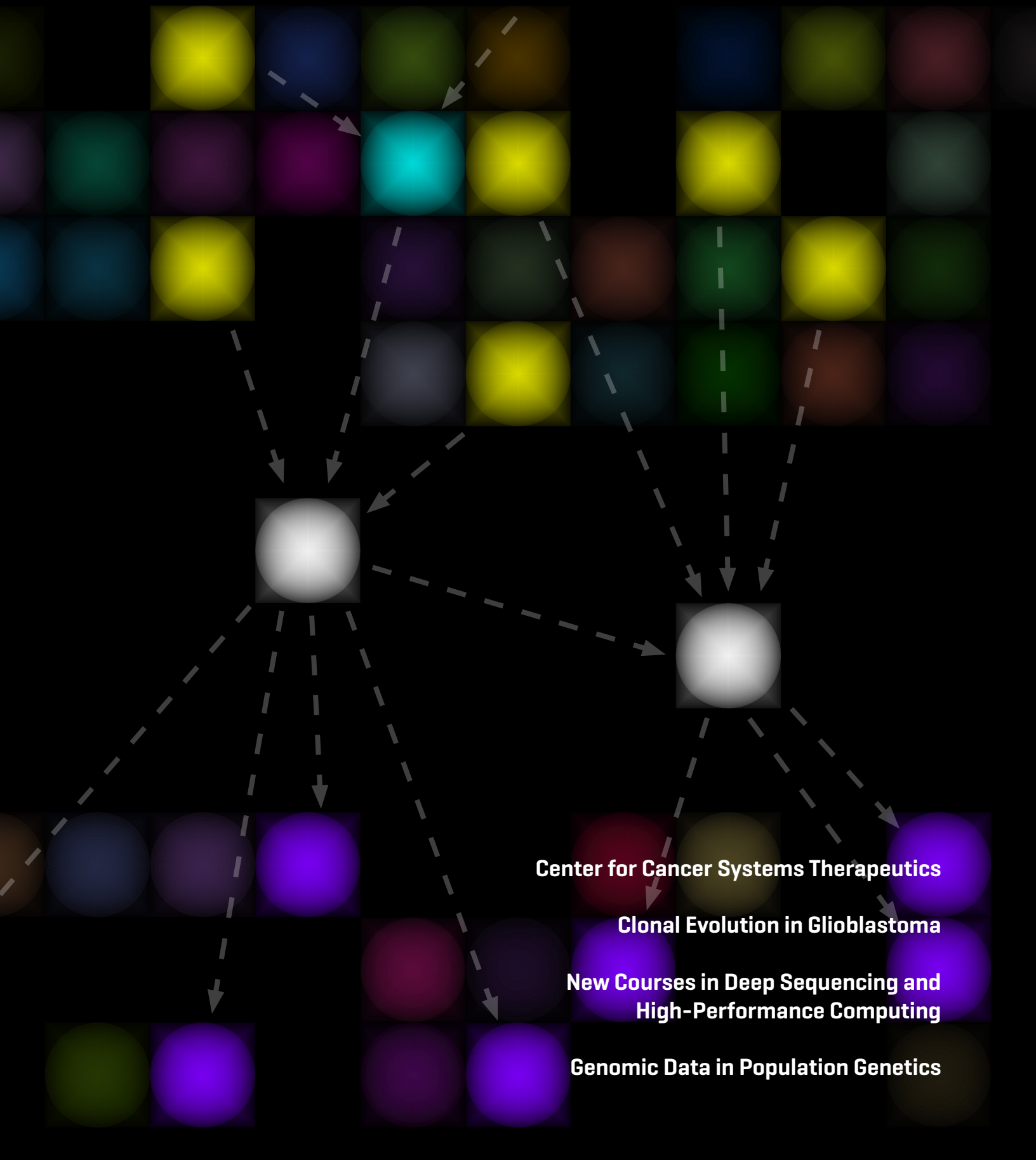
COLUMBIA UNIVERSITY
Department of Systems Biology

Research Highlights
September 2016



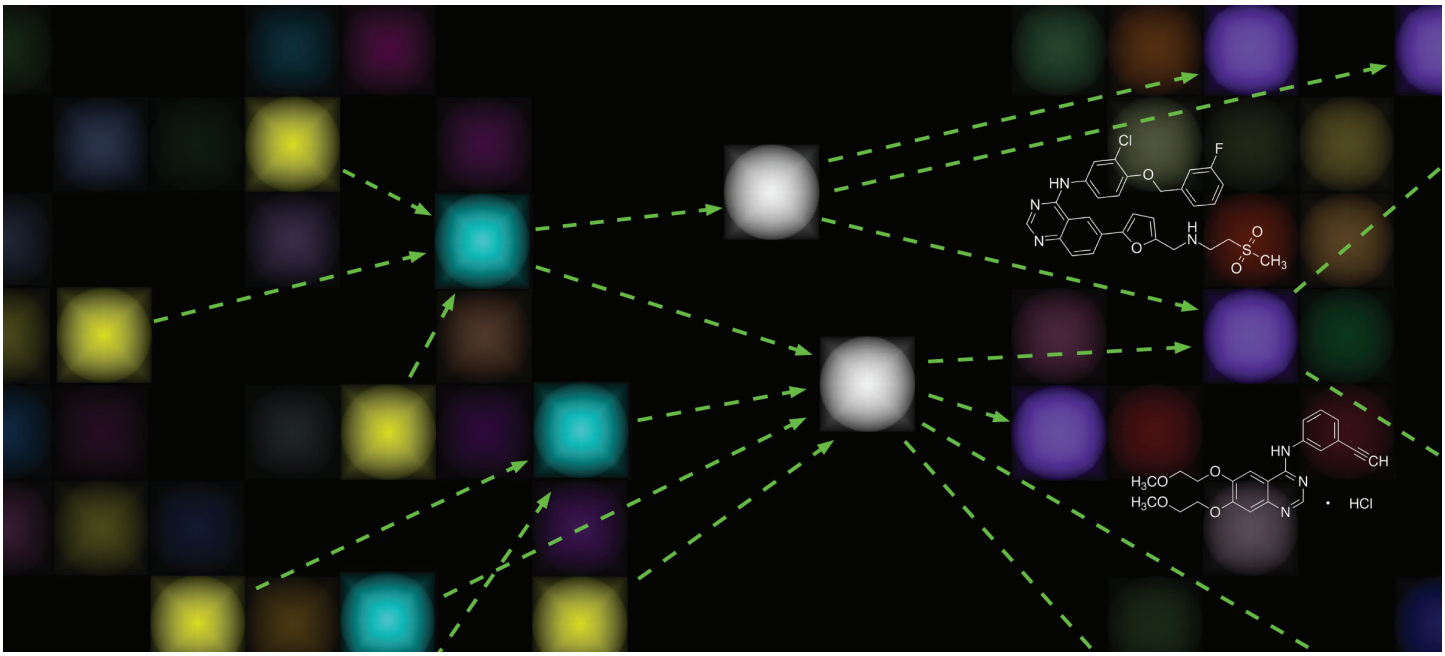
Center for Cancer Systems Therapeutics

Clonal Evolution in Glioblastoma

New Courses in Deep Sequencing and High-Performance Computing

Genomic Data in Population Genetics

Columbia Awarded NCI Center for Cancer Systems Biology



Master regulators of tumor homeostasis (white) integrate upstream genetic and epigenetic events (yellow) and regulate downstream genes (purple) responsible for implementing cancer programs such as proliferation and migration. CaST aims to develop systematic methods for identifying drugs capable of disrupting master regulator activity, considering how tumors evolve over time.

The Columbia University Department of Systems Biology has been named one of four inaugural centers in the National Cancer Institute's (NCI) new Cancer Systems Biology Consortium. This five-year grant will support the creation of the Columbia Center for Cancer Systems Therapeutics (CaST), a collaborative research center that will investigate the general principles and functional mechanisms that enable malignant tumors to grow, evade treatment, induce disease progression, and develop drug resistance. Ultimately, the Center plans to develop new methods for identifying drug treatments capable of targeting master regulators in the programs driving cancer homeostasis.

CaST will build on previous accomplishments in the Department of Systems Biology and its Center for Multiscale Analysis of Genomic and Cellular Networks (MAGNet), using systems biology methods to study the complex molecular machinery underlying cancer. At the same time, however, the new center aims to move beyond a static understanding of cancer biology toward a framework that can account for the dynamic, ever-changing nature of the disease. This more nuanced understanding could eventually enable scientists to better predict how individual tumors will change over time and in response to treatment.

Investigating cancer in this way means sidestepping the "oncogene addiction" paradigm, which has focused on identifying and targeting individual genetic alterations that are critical for the survival of cancer cells. Although this approach has led to some notable successes (for example, imatinib, erlotinib, and herceptin), currently a small minority of tumors seen in the clinic manifest actionable mutations. Moreover, targeted inhibitors have by and large shown limited long-term effectiveness, typically failing as tumors relapse and become drug resistant. For this reason, identifying new categories of druggable targets and finding better ways to manage tumors as they evolve constitute important challenges for the future of cancer research.

Instead of focusing exclusively on actionable mutations, CaST will concentrate on the regulatory machinery that enables established tumors to survive and grow. Driving its research is the hypothesis that cancer cells achieve and retain their malignant state by hijacking the regulatory programs responsible for normal cell development. Just as normal cells differentiate into specific cell types, tumor cells undergo a process of canalization in which molecular modules (tumor checkpoints) and specific proteins within these modules (master regulators) act like funnels, channeling the effects of upstream genetic alterations into cellular behaviors that are typical of cancer.

Such proteins, which appear consistently across many tumors, are typically not themselves the result of mutated genes. Nevertheless, they become essential for maintaining "tumor homeostasis" — the ability of cancer cells to survive and grow in a stable state. Previous work at Columbia has shown that master regulators are a type of "Achilles heel" for tumors, making them attractive targets for disrupting tumor homeostasis and cancer-related cell state transitions.

An additional component of CaST's efforts will be to address the question of how the heterogeneity seen in tumors affects tumor homeostasis and canalization. When subjected to anticancer drugs, this heterogeneity within a single tumor goes through a process of evolutionary selection, with different populations of cells reacting in different ways based on how their regulatory machinery processes the attack. The results of such selection pressures are often the downfall of existing therapies, as a portion of the tumor evades treatment and continues to grow.

Commonly used methods for profiling tumor tissue, including bulk sampling and sequencing techniques that generate an average profile of a tumor, have been incapable of distinguishing this fine-grained structure. For this reason, CaST will develop and test new methods for differentiating diverse cell populations within individual tumors. These tools will include new high-throughput single-cell sequencing technologies as well as patient-derived xenografts, in which human tumor tissue is implanted in mice to study how it grows and responds to perturbations.

Identifying tumor checkpoints and master regulators of specific tumor states, the scientists propose, could dramatically simplify the problem of prioritizing drugs and drug combinations against cancer. Using new high-throughput drug screening technologies developed in the Department of Systems Biology, the investigators plan to identify compounds whose mechanism of action can disrupt specific master regulator proteins responsible for tumor homeostasis and state transitions.

"We are very excited to have the opportunity to explore how systems biology approaches could help address some of the most critical questions facing precision cancer medicine today," says Andrea Califano, chair of the Department of Systems Biology and a co-principal investigator for the center. "By assembling this multidisciplinary team of very dynamic scientists, we think CaST should be in a unique position to push the field forward."

The award is particularly notable for the Department of Systems Biology because it now has grants from all three of the NCI's programs supporting research centers in this field. Other Department of Systems Biology national centers are funded through the Cancer Target Discovery & Development Program and the Physical Science-Oncology Program.

For more information about the Center for Cancer Systems Therapeutics, visit **systemsbiology.columbia.edu/cast**.

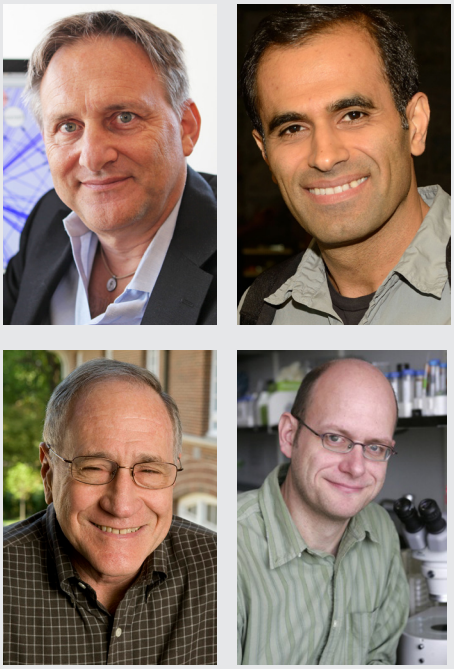
Distinguished Awards for Department of Systems Biology Faculty

Andrea Califano, Chair of the Columbia University Department of Systems Biology, was named a recipient of a National Cancer Institute Outstanding Investigator Award. The seven-year grant will support the development of systematic approaches for identifying the molecular factors that lead to cancer progression and to the emergence of drug resistance at the single-cell level. In addition, he was named a AAAS Fellow in recognition of his contributions to advancing science.

Saeed Tavazoie received a 2015 National Institutes of Health Transformative Research Award. The grant will support research to develop state-of-the-art experimental and computational methods for comprehensively mapping and modeling all pairwise molecular interactions inside cells. The Transformative Research Award is a part of the NIH Common Fund's High-Risk, High-Reward Research program.

The International Society for Computational Biology elected **Barry Honig** to its 2016 ISCB Class of Fellows. The award recognizes distinguished ISCB members who have shown excellence in research and/or service to the computational biology community. Dr. Honig's award acknowledges his "seminal contributions to protein structure prediction and molecular electrostatics, and his more recent work on protein function prediction, protein-DNA recognition, and cell-cell adhesion."

Oliver Hobert, an interdisciplinary faculty member of the Department of Systems Biology, has received a Javits Neuroscience Investigator Award from the National Institute of Neurological Disorders and Stroke (NINDS). This prestigious grant will enable investigation of sex-based differences in the regulation of neuronal identity.

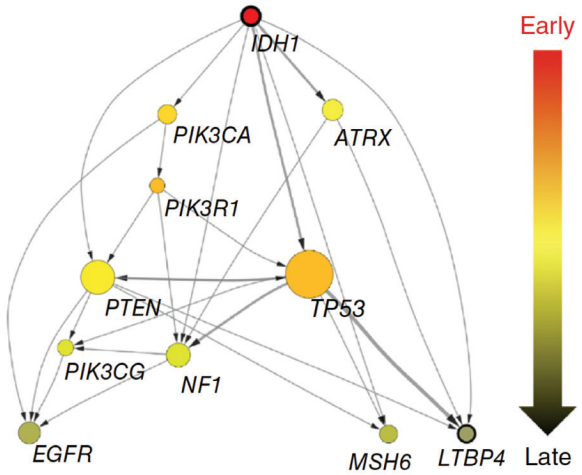


Clockwise from top left: Andrea Califano, Saeed Tavazoie, Oliver Hobert, and Barry Honig.

Glioblastoma Tumor Evolution and Strategies against Advanced Disease

Glioblastoma multiforme (GBM) is the most common and most aggressive type of primary brain tumor in adults. Existing treatments against the disease are very limited in their effectiveness, meaning that in most patients tumors recur within a year. Once GBM returns, no beneficial therapeutics currently exist and prognosis is generally very poor.

To better understand how GBM evades treatment, an international team led by Antonio Iavarone and Raul Rabadan at the Columbia University Center for Topology of Cancer Evolution and Heterogeneity has been studying how the cellular composition of GBM tumors changes over the course of therapy. In a paper published in *Nature Genetics*, they provide the first sketch of the main routes of GBM tumor evolution during treatment, showing that different cellular clones within a tumor become dominant within specific tumor states. The study uncovers important general principles of tumor evolution, novel genetic markers of disease progression, and new potential therapeutic targets.



Schematic showing the order in which driving mutations change as glioblastoma tumors progress.

The investigators' approach is based on a growing appreciation of the importance of clonal heterogeneity within tumors to their response to treatment. This perspective suggests that individual tumor cells undergo a process of Darwinian selection when subjected to therapy. Certain cells are eliminated, but a subset of the ones that survive are resistant and continue to multiply along specific lineages, becoming the ultimate cause of death. Despite this general conceptual understanding, however, scientists have found it very challenging to identify consistent patterns of clonal evolution in solid tumors that would enable them to design specific therapies for specific disease states.

The scientists hypothesized that if tumors evolve along clonal lineages made up of subpopulations of related cells, genomic profiling of cells before treatment and at the time of disease recurrence might enable them to develop the equivalent of family trees of tumor cell ancestry. They performed longitudinal genomic and transcriptomic analyses of tumors from 114 patients with GBM, using this data to reconstruct the evolutionary history of each tumor. Initially, this enabled them to identify mutations, copy number variations, expression profiles, and gene fusions that are differentially associated with specific disease states, including several genetic alterations that had never before been identified.

Using mathematical approaches developed in the Rabadan Lab, the team then used information about changes in the clonal composition of the many individual tumors to build an overarching model of GBM evolution. Surprisingly, the researchers discovered that tumor evolution over the course of treatment is not a linear process. Although one might expect that the cells present after treatment would be descendants of those present before treatment, they instead found that the dominant clones after treatment lack many of the mutations seen in dominant clones before the introduction of therapy. In fact, their model revealed that in most patients the clones responsible for driving different disease states must have diverged from a common ancestral cell many years before the tumor was detected.

This finding suggests that GBM tumors undergo clonal replacement over the course of treatment, with different lineages independently driving an aggressive tumor phenotype at different disease stages. Using techniques the Rabadan Lab developed previously and initially used in a study of leukemias, the investigators constructed what they call a tumor evolutionary directed graph. The model revealed that mutations in *IDH1*, *PIK3CA*, and *ATRX* occur early in GBM; mutations in *TP53*, *NF1*, and *PTEN* occur later; mutations in EGFR are subclonal and occur late, mostly at the time of diagnosis; and mutations in *MSH6* and *LTBP4* occur specifically during relapse.

Distinguishing alterations that are common to all cells within a tumor versus those that are specific to diagnosis or recurrence could provide valuable information for optimizing targeted therapeutic strategies and clinical trials. For instance, the analysis suggests that EGFRvIII inhibitors, which have been under investigation for treatment of glioblastoma, might not be an ideal choice for targeted therapy because the variant disappears as the tumor evolves. It is more likely that a therapy targeting pre-existing clonal populations that are likely to drive advanced, resistant disease — even if these cell populations are small at the time of diagnosis — could be successful.

"Genetic alterations such as mutations and gene fusions that are present in most cells in untreated glioblastoma tumors and persist after treatment with standard-of-care therapies should provide

clinicians with the best possible targets for precision medicine," Dr. Iavarone explains.

The finding about *LTBP4*'s association with disease recurrence is particularly encouraging, the paper reports, because the corresponding protein is known to regulate the TGF-β pathway, which has previously been shown to be important in aggressive gliomas. In a series of laboratory experiments using glioma cell lines, they showed that silencing *LTBP4* markedly reduced cell proliferation. The scientists are currently investigating whether this discovery could offer an effective strategy for eliminating the clones that drive advanced GBM.

In addition, they determined that methylation of the gene *MGMT* at diagnosis predicted longer survival, while low expression of *MGMT* at recurrence was significantly related to better prognosis. Most interestingly, the team found relapse-specific gene fusions involving the *MGMT* gene that drive resistance to treatment. Such a diagnostic marker could in the future potentially serve as a prognostic marker and help clinicians in treatment planning.

"This work illuminates the power of genomic techniques for precision medicine approaches to cancer," Rabadan says. "Despite the complexity of these tumors, we were able to identify mutations in pathways associated with progression, new fusion genes contributing to drug resistance, and novel markers of survival. We think that this opens some promising avenues for treating aggressive GBM tumors, and for implementing detection techniques that could help in early diagnosis."

The research was conducted as part of Columbia University's Center for Topology of Cancer Evolution and Heterogeneity, one of five national centers in the National Cancer Institute's Physical Sciences–Oncology Network. Other scientists who participated in this study are based at the Fondazione IRCCS Istituto Neurologico Besta (Milan, Italy), Samsung Medical Center (Seoul, South Korea), and the University of Texas at Austin.

Related publication
Wang J, Cazzato E, Ladewig E, et al. **Clonal evolution of glioblastoma under therapy**. Nat Genet. 2016 Jul;48(7):768-76.

Center for Topology of Cancer Evolution and Heterogeneity Forms Discussion Groups and Issues First Startup Grants

Opened in 2015 with the support of the National Cancer Institute's Physical Sciences-Oncology Program, the Center for Topology of Cancer Evolution and Heterogeneity is an interdisciplinary center based at Columbia University Medical Center (CUMC) that is studying how subclonal populations of cells evolve in solid tumors. As part of its education and outreach efforts, the Center initiated two discussion groups as well as a program to teach students how math and physics can support genomics research.

The New York Metropolitan Area Discussion Group in Mathematics and Oncology held three events in 2015-2016, with speakers including Harold Varmus, Gunnar Carlsson, Arnold Levine, Tom Maniatis, and others meeting at CUMC, New York University, and the Institute for Advanced Study. A second discussion group, called the New York Applied Topology Meetings, met biweekly, bringing together researchers with interests in applied topology. The program included an eight-week introduction to topological data analysis (TDA) by Princeton University mathematician Michael Lesnick, as well as talks by researchers at Columbia and other institutions who are using TDA in various biological contexts. Education and outreach efforts at the Center were also enhanced with the launch of the PSOC Summer Immersion Program (PSIP), which offers students from diverse backgrounds, including from Columbia University and international institutions, the chance to become immersed in the Center's research.

The Center also started a pilot grant program to support interdisciplinary research. Pilot projects are enabling the development and testing of new mathematical approaches within the context of cancer research, giving mathematicians and physicists experience working in biological settings, and providing cancer biologists opportunities to explore how mathematical methods can be used to guide their research.

The coming year's activities will include East Coast meetings held in partnership with the Dana-Farber Cancer Institute, Harvard University, and MIT.

To learn more, visit **psoc.c2b2.columbia.edu**.

How Genomic Data Are Changing Population Genetics

Common chimpanzee at the Leipzig Zoo. Photo: Thomas Lersch, Wikimedia Commons.

Molly Przeworski, a professor in the Columbia University Departments of Biological Sciences and Systems Biology, majored in mathematics at Princeton before beginning her PhD in evolutionary biology at the University of Chicago in the mid-1990s. While there, she realized that the availability of increasingly large data sets was changing population genetics, and has since been interested in using statistical approaches to investigate questions such as how genetic variation drives adaptation and why mutation rate and recombination rate differ among species. In the following interview, she describes how population genetics is itself evolving, as well as some of her laboratory's contributions to the field.

Population genetics is a discipline that has been changing over the past couple of decades. Could you talk about how you have seen it develop in your career so far?

Since the 1920s, population genetics has been focused on modeling evolutionary processes that occur on time scales that are too slow to be observed. In its early days, even the object of study—genetic variation among individuals—was very hard to measure. The first genetic loci that were found to be variable in humans were those responsible for determining blood groups, because they were easy to assay. But those kinds of variable loci—or "polymorphisms"—were few and far between. The vast majority of genetic variation among individuals was completely inaccessible until the 1960s, when people started developing techniques to look at protein variants, and until after 1983, when Marty Kreitman and others started using sequencing to survey genetic variation in populations. For many decades, then, it was a strange field, which was trying to ask deep questions about evolutionary processes without having access to the data that were needed for inferring what might have happened in the past.

That all changed in the late 1990s. As improvements in technologies for genome sequencing made it less labor intensive, a trickle of data started coming in. By the time I completed my degree it was clear that more were on the way and that statistical approaches would be needed to analyze them. Later, during a postdoc with Peter Donnelly in the Oxford University Department of Statistics, I became interested in the idea that it should be possible to learn not just the molecular basis of human adaptations, but when they occurred in our evolutionary history. I developed a statistical method to tackle this problem.

Around this time I met Svante Pääbo, who later became best known for sequencing the Neanderthal genome, but had just determined that the gene *FOXP2* had been under natural selection at some point in human evolution. The gene is of particular interest

because it plays a role in speech and language development in humans. His lab had shown evidence that the gene was involved in some kind of adaptation in human ancestors, but he didn't how to date when it occurred. He invited me to work at the Max Planck Institute in Leipzig. I did this both because I was interested in this specific application and because it got me closer to real data, which I had been chasing at the time.

In the years since then, thousands of genomes of humans and most organisms you can think of have become available, and population genetics has become a field with almost unlimited data. This has made it possible to ask many new kinds of questions and finally put century-old theories to the test. In my current work I focus on adaptation and the processes that generate genetic variation.

How do statistical and computational methods help to explore these kinds of issues?

I've always been interested in questions related to the origins of genetic differences among humans and other species: What fraction of those differences confers fitness advantages or is there by chance? What are the processes that produce genetic variation? And how does that variation play out through population dynamics to bring about adaptations such as bipedalism in humans or eyespots on butterflies?

To conceive of how we investigate this, think of your ancestry. You have many ancestors in your family tree, but in any particular position of your genome, you only inherit DNA from two of them. Which of the two ancestors you inherit DNA from changes across positions in the genome because of recombination, the shuffling of segments of DNA that is a natural part of cell division. Over long time scales, this means that if you compare your chromosomes to someone else's, you might have close ancestors in common for some bits of your genome, but not for others.

Now imagine that a beneficial mutation arises in somebody. Carriers of that beneficial mutation leave more offspring, and so it spreads through the population faster than variants that have a bad effect or no effect. Using computational means to analyze genomic data, we can observe genetic variation and identify regions of the genome where all individuals in our cohorts are very similar — unusually similar. We might look at hundreds of individuals, asking how many bits of DNA they have in common in any given position of the genome. These regions of high similarity indicate variants that spread very rapidly, and give us a hint that something in those regions was beneficial in the context of natural selection. We want to know how that process occurs, how long it takes, how strongly beneficial it was, and how many genetic alterations it takes to bring about complicated adaptations. In a sense, we're using experimentally obtained genomic data and statistical methods to reverse engineer the processes that drive evolution.

In addition to using computational methods to identify genetic variation, would you say that you are trying to understand the underlying genomic machinery that actually drives this variation?

Yes, I think that's right. The sources of all genetic variation, whether it leads to a disease or an adaptation, are chance changes to the genome. These can be mutation events or alterations that occur during recombination. For a long time, I've been interested in viewing these phenomena not just as inputs that produce other traits, but as traits that are themselves specified genetically. For example, the repair factors that determine how frequently mutations occur in a particular organism or that specify where in the genome the machinery for recombination should place itself — called recombination hotspots — are themselves genetically specified. This means that evolution acts on the very inputs to the processes that enable adaptation and natural selection.

This perspective raises several questions. For example, the mutation rate across species tends to be low but not zero; why is that? Or why does recombination occur in genes in yeast but more often outside of genes in humans? What are the evolutionary consequences? In this sense, I'm interested in understanding not just the mechanisms that produce genetic variation, but also why those values and patterns exist, and why they differ between species.

You mentioned you are interested in understanding how genetic variation enables adaptation. Can you give an example of your work on this topic?

One area that we're interested in exploring is the benefits of persistent genetic variation. If you consider variation in things like eye color, for example, variation itself can be tens of thousands to hundreds of thousands of years old; in exceptional cases it might be a million years old. In a paper we published a couple of years ago, we were interested in trying to see whether there are regions of the genome where the variation we see among humans is unusually old, where selection led to the maintenance of diversity in the population rather than one type outcompeting another.

We took a set of 120 human samples and 10 chimpanzee samples, and asked whether there are regions of the genome in which stability in variation among humans is so old that it predates the evolutionary split with chimpanzees. We had previously shown that the A and B blood groups are extremely old variations. They're millions of years old, so old that humans and gibbons have the same blood types because they inherited variants that were already present in their common ancestor. We were looking for similar regions in chimpanzees and found dozens of DNA segments where the same haplotypes are present in both humans and chimpanzees.

When identical, cross-species variation is so old, you know it's not there by chance, because if it were it would have been lost over the ages due to genetic drift. There must be some form of selection that makes it advantageous to maintain variation in the population, and we wanted to understand how this happens.

We found that many of the regions that humans and chimps share are involved in the production of membrane glycoproteins, a class of proteins that viruses use to enter host cells and that some bacteria imitate to evade being attacked by the host immune system. This finding led us to hypothesize that some of the examples we found of persistent variation could be related to pathogen-host coevolution. Typically, as resistance to a pathogen builds up in the human population, the pathogen has less of a foothold in its hosts. But as the pathogen evolves, the frequency of resistance decreases and then the pathogen shores itself back up. Over time you see a cycle that maintains both resistant and nonresistant pathogen types in the population. We think that the long-term, cross-species persistence of variation in the genes that mediate these relationships could reflect this.

Recently, a paper was published in which the authors conducted a genome-wide association study to map the genetic basis of susceptibility to malaria in Africa. Interestingly, the one significant hit of a genetic variant that modulates whether a person is susceptible to malaria or not was on one of the regions that we had found in our earlier paper as being unusually old in humans. We were happy to see this, because their discovery was consistent with the notion that variability is maintained in genomic regions involved in host-pathogen coevolution.

Modern population genetics clearly has a role in explaining features of evolution and natural selection, but are there ways in which it intersects with other biological fields?

In a lot of molecular biology or cell biology, an experiment involves breaking a system in some way — perhaps by introducing a mutation or doing a mutant screen — and then learning the function of the particular gene or pathway you broke by observing the experiment's effects. But when you study genetic variation using population genetics, you're looking at the results of a huge mutagenesis experiment that has already been performed. All of us are essentially mutants, and we all show differences in the activity of particular genes and in fluxes in different molecular signaling pathways. We're like a living molecular biology experiment, with the difference that you know that the individuals you are studying can survive with whatever mutations they have. Species lose genetic changes that are hugely deleterious through natural selection, so population genetics allows you to computationally try out all combinations of variants that an organism can have and still survive. This provides a tremendous amount of information to be mined about all kinds of genetic processes and how they work.

Molly Przeworski

From this perspective, one thing that's exciting about the analysis of genetic variation today is the possibility of learning about biology in a much broader set of organisms. The reason biologists study mice and fruit flies is that well-established genetic tools exist for these systems, and they can easily be bred. But instead of engineering mice with specific defects, it is now becoming possible to look at different individuals and use computational approaches to identify genetic factors that differentiate them with respect to particular traits. In this way, statistical approaches based in population genetics also allow you to learn a lot about the genetics of non-model organisms, because you don't actually have to do the breeding and knockout experiments that are necessary for traditional animal models.

In a recent paper our lab published in *Science,* we studied zebra finches. The species has been used as a model organism for learning, but in terms of genetic resources not much exists. By using statistical approaches for analyzing genomic data, we showed that we can learn a great deal about meiotic recombination in birds. What I'm particularly excited about these days is taking a very broad taxonomic perspective, beyond the handful of model organisms, to ask how recombination works in animals like fish, frogs, snakes, and turtles, and get a sense of the whole breadth of what is possible — and what's not possible — and how and why it changes over time.

Does population genetics have implications for human health?

Population genetics has many implications for disease genetics, specifically for understanding genetic susceptibility to complex diseases. Population genetics aims to interpret genetic variation, and if we want to predict who is more susceptible to disease than someone else, map the relevant genetic variants, and distinguish the roles of environment and genetics in causing a disease, these

are all traditional population genetics questions. In some sense, human genetics today is applied population genetics: looking at variation among humans and relating it to variation in phenotypes.

There are many other areas of synergy between population genetics and other areas in biology: for example, my lab works on variation in recombination, which is a key to understanding aneuploidy (i.e., when fetuses have the wrong number of chromosomes), a major source of infertility in humans. It turns out that the major mechanism by which this occurs involves errors in recombination, and so understanding what is tolerable and what is not tolerable in terms of variation in recombination is critical to understanding the genesis of this phenomenon.

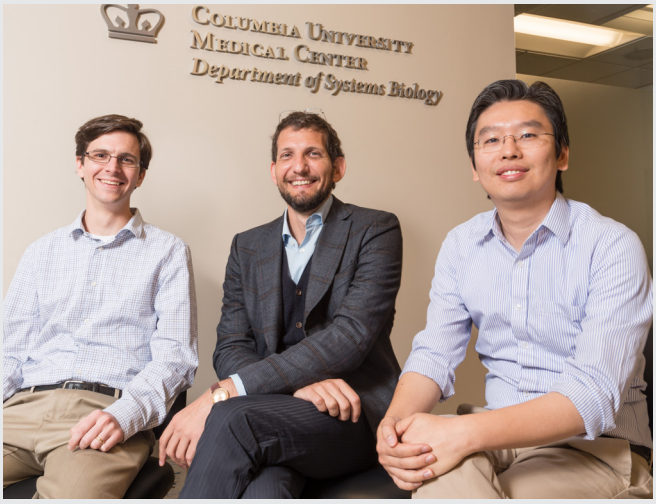
With all of the data you now have to work with, it sounds like it's a golden age for population genetics. Do you see other developments on the horizon?

What's exciting is not just that there are a lot of data, but also that it is now possible to test the many evolutionary theories that population genetics has produced. Considering that it's only recently that we have this kind of genomic data, it's fascinating to me that Darwin in particular turned out to be right about almost everything he wrote. The level of detail in the evidence we can now marshal is mind-boggling, and it all supports his theory. I also continue to find it remarkable how useful the models developed by earlier population geneticists have turned out to be for making sense of the data people are generating today. Their continued relevance speaks to the power of abstract reasoning.

For 30 years the evolutionary literature has posited, for example, that if you see conservation of a particular bit of DNA in distantly related species, it must be really important. Now people are using this concept when they scan for mutations that cause disease. If you perform genome sequencing on a patient and notice that there were two mutations in a gene, you might ask whether this is unusual. Information about whether that gene is evolutionarily conserved and whether mutations in it can be tolerated can therefore be hugely informative, as it can help to prioritize mutations that are most likely to cause disease. Evolutionary principles that were developed long before these kinds of data became available turn out to be really powerful.

My sense is that there is an increasing synergy between many areas of biology and human genetics. In some ways we might be seeing the end of population genetics, as analysis of genetic variation becomes an essential part of molecular biology, cell biology, neuroscience, and other fields. Over time, a lot of tools developed within the context of population genetics are likely to be adopted by those disciplines. Perhaps they will even subsume population genetics, while evolutionary biology will keep asking more specific questions about evolution and natural selection.

Department of Systems Biology Opens New Biotechnology Hub



Assistant Professors Peter Sims, Sagi Shapira, and Harris Wang recently moved into lab space in Lasker. Photo: Lynn Saville.

The Columbia University Department of Systems Biology has opened a new experimental research hub focused on biotechnology development. Occupying one and a half floors in the Mary Woodard Lasker Biomedical Research Building, the facility brings technology-focused investigators into close proximity, with the goal of facilitating collaborations that will lead to the development of new methods for the study and engineering of biological systems.

"New technologies give you access to fundamentally new ways of looking at biological processes," says Harris Wang, a synthetic biologist whose lab relocated to Lasker. "By the time there's a commercial kit to do something, a lot of the interesting questions have already been answered. The research communities that are going to be most successful are always the ones that are creating new applications."

Moving into a 2,000 square-foot laboratory space has also enabled a substantial expansion of the JP Sulzberger Columbia Genome Center's next-generation sequencing capabilities, including the potential to add larger instruments for sequencing genomes of large patient cohorts. The Genome Center's move reflects the steady growth in demand for high-quality genomic data over the past several years, particularly at the Herbert Irving Comprehensive Cancer Center.

Method for Determining Protein Function Opens Doors for Precision Medicine

In a paper in *Nature Genetics*, the laboratory of Andrea Califano introduces what it describes as the first method for analyzing a single tumor biopsy to systematically identify proteins that drive cancerous activity in individual patients. Using knowledge gained by modeling networks of molecular interactions, their computational algorithm, called VIPER (Virtual Inference of Protein activity by Enriched Regulon analysis), offers a unique new strategy for understanding how cancer cells survive and for identifying personalized cancer therapeutics.

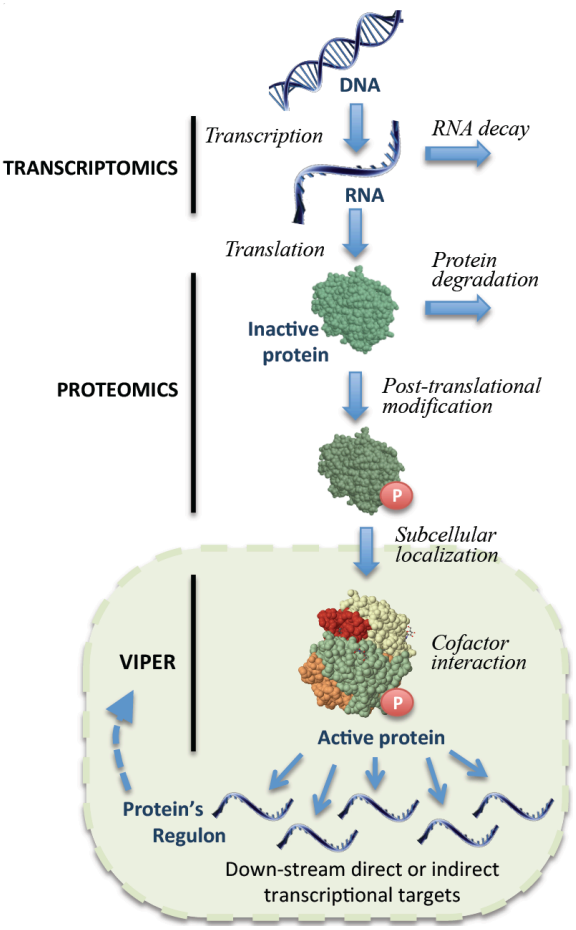
According to Dr. Califano, "VIPER makes it possible to find actionable proteins in 100% of cancer patients, independent of their genetic mutations. It also enables us to track tumors as they progress or relapse to determine the most appropriate therapeutic approach at different points in the evolution of disease. So far, this method is looking extremely promising, and we are excited about its potential benefits in finding novel therapeutic strategies to treat cancer patients."

At its core, cancer is a constellation of diseases that arise when normal protein activity in cells goes awry, causing them to grow and spread uncontrollably. Because proteins are the products of genes, cancer biologists have for many years hypothesized that tumor cells become addicted to the mutated oncogenes that are responsible for the initial tumor growth. By identifying and targeting the proteins harboring these mutations, the reasoning goes, it should be possible to design personalized therapies that could halt cancer.

Research in this direction has led to some important successes. In general, however, it has thus far improved treatment for only a small minority of tumors. In part, this longstanding focus on genes as a proxy for protein activity has been a consequence of the strengths and limitations of available technologies. Next-generation DNA sequencing is consistently reproducible, a factor that is essential for clinical applications. However, gene sequence alone cannot reveal whether the corresponding protein is actually aberrantly activated. This is because proteins operate in cooperation, collectively forming complex interaction networks that influence and ultimately determine whether a mutation will affect a cell's behavior.

Even when cancer-driving mutations have been identified, researchers have faced an uphill battle in identifying durably effective therapeutics. One reason is that tumors typically develop drug resistance as they evolve. Another is that some of the most commonly recurring mutations block the activity of genes called tumor suppressors, such as *TP53* and *PTEN*. Since drugs generally work by inhibiting proteins — not activating them — mutated tumor suppressor proteins are not good therapeutic targets.

From this perspective, being able to identify cancer-driving oncoproteins through direct measurement of protein activity across the course of disease would be more desirable. But although current technologies such as mass spectrometry can measure protein abundance, they are too expensive and complicated for use in clinical applications, and cannot systematically account for a wide range of factors — such as post-translational modifications or protein localization in specific parts of a cell — that affect protein activity. New methods for identifying the proteins responsible for driving cancer would therefore be a valuable addition in the fight against disease.



By analyzing downstream gene expression patterns, VIPER can identify proteins that are critical to cancer cell survival.

The Califano Lab's approach to identifying oncoproteins is based on past work revealing that although cancer cells can harbor an extremely heterogeneous repertoire of genetic alterations, these mutations enable them to misuse the complex network of molecular interactions that regulate their behavior in extremely similar ways. Even though many genetic mutations are present in tumor cells, Califano has shown that these mutations converge to and

are integrated by specific proteins called master regulators, which activate the programs that are necessary to make a cancer cell. Importantly, these proteins are not themselves mutated, but are nevertheless essential for maintaining cells in their cancer-related state. Previous work has also indicated that these master regulators are relatively few and are conserved across a large subset of cancer patients. Identifying them and finding ways to target them could thus simplify the landscape of cancer dramatically, especially when compared to the myriad ways in which a tumor cell's genome can be mutated.

Developed by research scientist Mariano Alvarez, VIPER is designed to do so based on a very simple concept. Rather than measuring the activity of a protein directly, it infers activity based on the expression of the genes the protein regulates. The Califano Lab first uses ARACNe, an algorithm that has been broadly adopted and validated by the research community, to identify targets of all proteins in a specific tumor type. By applying a novel statistical framework for analyzing gene expression data, VIPER then determines the activity of all cancer-relevant proteins, identifying those that are abnormally activated in a specific patient's tumor.

"It's like detective work to determine which of two crime families was the mastermind behind a murder," Califano explains. "First you build a map of the two organizations and then look for fingerprints or eyewitness accounts of who was at the crime scene. If you identify someone who is a part of one of the two organizations, you can quickly figure out the head of the organization who gave the order. We can understand protein activity in a similar way by observing expression changes in the genes they regulate."

Importantly, the researchers report that because VIPER measures each protein's activity based on the expression of hundreds of genes, their measurements are reproducible and thus appropriate for potential use in the clinic. This is the case even though individual gene expression measurements may not be reproducible. This feature makes it possible to investigate protein activity using formalin-fixed, paraffin-embedded (FFPE) tissue samples, which are clinically more common than fresh tissue samples but are typically degraded in ways that have in the past made them difficult to analyze.

In addition, drugs are already known to be capable of targeting many of the specific proteins that VIPER identifies. And because RNA sequencing (RNASeq) is less expensive than genome sequencing and VIPER requires just a single tissue sample, this approach is a tenth of the cost of genetic sequencing. This makes it feasible to use it repeatedly during the course of a patient's cancer treatment. For example, if a tumor stops responding to a particular therapy, a new VIPER analysis could be performed to determine how it has evolved and which new druggable proteins are now essential for its survival.

As the paper reports, the Califano Lab performed a number of studies to validate VIPER's effectiveness. In one they analyzed 173 basal breast carcinomas recorded in the Cancer Genome Atlas (TCGA). Even though these samples were ostensibly the same type of cancer, the investigators found that their gene expression patterns were wildly different from sample to sample, making it very difficult to predict a therapy that might work for all patients from expression alone. However, when they used the network-based approach that VIPER offers, they discovered that essential cancer-driving proteins were consistently present across all samples. Such findings are exciting in light of other work in the Califano Lab that has indicated that these proteins — which are different from those produced by typical oncogenes like *BRAF*, *EGFR*, and *ERBB2* — could be important tumor checkpoints across many cancer types.

"This makes it possible to use a more universal treatment for all cancers in this subtype," Califano explains. "If there are drugs that can target those proteins, you don't need to figure out how each one of those cancers is different at the genetic level."

In the end, the scientists used VIPER to investigate more than 10,000 tumor samples representing 14 different malignancies from the TCGA repository. Their findings indicate that VIPER could identify dysregulation in cancer cells that was the result of mutations, as well as proteins whose abnormal activity did not arise from mutations in their corresponding genes but gained their cancer-driving ability because of other alterations. This suggests that available drugs could be effectively used in a substantial subset of patients who do not harbor mutations in their target genes.

VIPER is quickly becoming an important tool in Columbia's precision medicine initiative, complementing and extending other key strategies, including immunotherapy and genomic medicine. In a series of N-of-1 cancer clinical trials, the Califano Lab is working with clinical researchers in CUMC's Herbert Irving Comprehensive Cancer Center, using VIPER to analyze tumor samples from individual patients, identify proteins that are driving cancerous activity, and connect them to existing FDA-approved and investigative drugs that are already known to be able to target them. Although directly providing treatment is beyond the scope of the trials, findings from these studies have already enabled Columbia oncologists to recommend therapies that have extended survival and improved quality of life in patients.

Related publication

Alvarez MJ, Shen Y, Giorgi FM, et al. **Functional characterization of somatic mutations in cancer using network-based inference of protein activity.** Nat Genet. 2016 Aug;48(8):838-47.

Identifying Genetic Alterations that Modulate Gene Expression

A team of Columbia University researchers led by Harmen Bussemaker has proposed a novel approach for discovering some critical components of gene regulatory networks. Using statistical methods to analyze biological data in a new way, the researchers identified genetic alterations they call connectivity quantitative trait loci (cQTLs), a class of variants in transcription cofactors that affect the connections between specific transcription factors (TFs) and their gene targets.

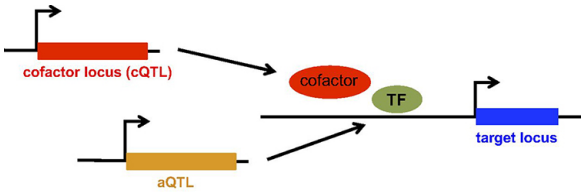
The work grew from the Bussemaker Lab's longstanding interest in understanding how transcription factors regulate the transcription of genes into mRNA. Genetic alterations in a TF's binding sites in genomic DNA can alter its binding and lead to changes in gene expression. In some cases this can lead to disease. For this reason, many scientists are searching for such alterations in the genome, which they call expression quantitative trait loci (eQTLs).

In 2010 Bussemaker proposed the existence of a new type of QTL. These so-called transcription factor activity QTLs (aQTLs) do not affect the TF binding sites themselves, but rather influence how much of a TF protein in the cell is present in the nucleus, where the genomic DNA resides. More recently, in collaboration with researchers at the Netherlands Cancer Institute, his lab used an extension of the aQTL approach called locus expression signature analysis (LESA) to analyze how viral insertions cause cancer in mice by affecting the protein-level activity of specific TFs. They also applied the aQTL approach to post-transcriptional networks that regulate mRNA stability.

The new paper extends this line of thought further by considering the fact that a transcription factor does not work alone to control gene expression. Instead, it relies on interactions with cofactor gene products that enhance the strength of its binding to its target genes, or its ability to interact with proteins that help it activate or repress its gene targets. Bussemaker hypothesized that genetic alterations in those cofactors could change their interaction with their partner transcription factors, altering the strength of TF binding and thereby influencing downstream gene expression. This could in turn, for example, change the way in which the cell processes an incoming signal from a drug. The team calls these alterations cQTLs.

Running the algorithm requires data representing genome sequence, gene expression levels, and prior information about transcription factor binding preferences. It then integrates and analyzes the data in several steps using statistical approaches.

The scientists tested their method in two yeast strains, sampling approximately 100 known transcription factors and then focusing on seven whose functional connectivity with their target genes seemed to be affected by genetic alterations. One of these, called *Ste12p*, was previously known to be an activator of the mating response pathway in yeast in the presence of a small molecule called α-pheromone. They applied their algorithm to find cQTLs related to *Ste12p*, and then used existing protein-protein interaction data to identify protein products within these regions that are physically capable of interacting with the transcription factor as cofactors.



Cofactors work with transcription factors (TFs) to enable efficient transcription of a TF's target gene. Genetic alterations in the cofactor gene (cQTLs) change the nature of this interaction, affecting the connectivity between the TF and its target gene.

One of the regions they pinpointed included a gene called *DIG2*, which had previously been shown to be a regulator of *Ste12p* activity. To determine whether the naturally occurring genetic alteration in the DIG2 protein was indeed a cQTL for *Ste12p*, they engineered a strain that was previously susceptible to α-pheromone so that it now differed by a single nucleotide in *DIG2*. Stimulating the strain with the small molecule led to a dramatic drop in Ste12p-mediated gene expression. Conversely, they engineered another strain that was previously not responsive to α-pheromone so that it contained the active nucleotide sequence in the predicted cQTL, and saw that it subsequently displayed the *Ste12p* expression activity normally seen in the mating response. Being able to control gene expression in this way, along with other findings described in the paper, confirmed the ability of the researchers' algorithm to identify cQTLs accurately.

Although their method was applied to a humble yeast system, the authors indicate it could offer a way to identify cQTLs in any organism for which the necessary data are available. "So far, I like to think of this as personalized medicine for yeast," Bussemaker remarks. "But with the increasing availability of genotype and expression profile data in projects like the Genotype Tissue Expression (GTEx) project, and considering the recent arrival of high-quality in vitro protein-DNA interaction data, we think we should be able to apply these same methods to identify cQTLs in humans as well."

Related publication

Fazlollahi M, Muroff I, Lee E, et al. **Identifying genetic modulators of the connectivity between transcription factors and their transcriptional targets.** Proc Natl Acad Sci U S A. 2016 Mar 29;113(13):E1835-43.

New Courses in Deep Sequencing and High-Performance Computing

Chaolin Zhang led a discussion in a new course on experimental and computational aspects of deep sequencing. Photo: Lynn Saville.

As the cost of next-generation sequencing has fallen and access to high-performance computing platforms has grown, the kinds of questions that biologists can ask and the methods they have for answering them have been changing rapidly. This presents some challenges for graduate education, however, as it means young scientists must gain new kinds of expertise in order to participate in cutting-edge biological research. The Department of Systems Biology recently supported the creation of two new courses aimed at providing these skills.

The first, titled Deep Sequencing, was developed by assistant professors Yufeng Shen, Peter Sims, and Chaolin Zhang, and provides the basics in both the experimental and analytical dimensions of the discipline. Among the topics covered are the history and development of modern sequencing technologies, an introduction to foundational statistics and algorithms, laboratory and analysis techniques for whole genome and exome sequencing and their applications in medical genetics, and methods related to RNA-seq and the study of transcriptional and post-transcriptional regulation. The course also includes a focus on cancer genomics and insights into single-cell sequencing and analysis.

As Zhang explains, "What makes this course unique is how it combines several important aspects of deep sequencing. The three instructors have very different backgrounds, but each of us relies heavily on deep sequencing in our research and has first-hand experience of the most recent developments."

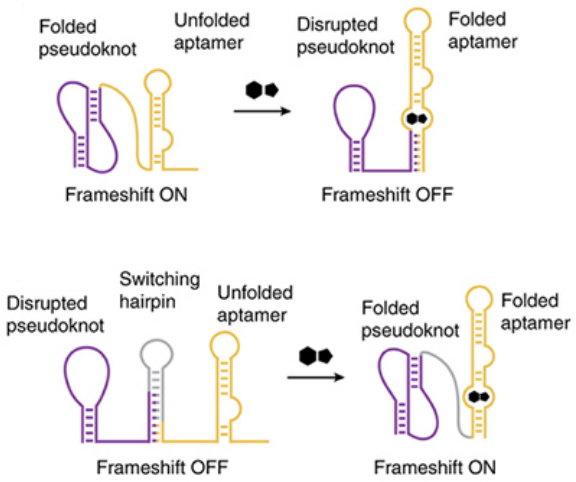
The second course, developed by Rebecca Yohannes, director of high-performance computing at the Mailman School of Public Health, and Hugh Ediet, lead engineer of the Department of Systems Biology's Information Technology Group (DSBIT), addresses practical challenges users face in programming high-performance computing clusters as well as theoretical questions that they raise. Topics being covered include parallel computing theory, as well as examples of the design, analysis, and implementation of high performance computing applications across a variety of scientific disciplines. Students also learn about high-performance computing system architecture and the basics of evaluating computing performance.

"Across the sciences," Yohannes says, "making sense of the large amounts of data out there requires high-performance computing, so these are critical skills. We're excited that the Mailman School was able to work with the Department of Systems Biology to design the first introductory course in supercomputing at Columbia, and the only supercomputing course among our peer institutions of public health."

Graduate Students Invent Technique for Reprogramming Translation

The ribosome is a reliable machine in the cell, precisely translating the nucleotide code carried by messenger RNAs (mRNAs) into the polypeptide chains that form proteins. But although the ribosome typically reads this code with uncanny accuracy, translation has some unusual quirks. One is a phenomenon called -1 programmed ribosomal frameshifting (-1 PRF), in which the ribosome begins reading an mRNA one nucleotide before it should. This hiccup bumps translation "out of frame," creating a different sequence of three-nucleotide-long codons. In essence, -1 PRF thus gives a single gene the unexpected ability to code for two completely different proteins.

Recently Andrew Anzalone, an MD/PhD student in the laboratory of Virginia Cornish, set out to explore whether he could take advantage of -1 PRF to engineer cells capable of producing alternate proteins. Together with Sakellarios Zairis, another MD/PhD student in the Columbia University Department of Systems Biology, the two developed a pipeline for identifying RNA motifs capable of producing this effect, as well as a method for rationally designing -1 PRF "switches." These switches, made up of carefully tuned strands of RNA bound to ligand-sensing aptamers, can react to the presence of a specific small molecule and reliably modulate the ratio in the production of two distinct proteins from a single mRNA. The technology, they anticipate, could offer a variety of exciting new applications for synthetic biology.



Two kinds of -1 PRF switches: In the first case, a pseudoknot that stimulates -1 PRF is energetically dominant, producing high frameshift levels. When a specific small molecule is present, the aptamer folds, disrupting the pseudoknot structure. In the second, a switching hairpin disrupts the pseudoknot and frameshifting. In the presence of ligand, the aptamer folds and destabilizes the switching hairpin, allowing the pseudoknot to induce a frameshift.

Although scientists had been aware of -1 PRF for some time, it was unclear when Anzalone began his project what specific features in RNA could cause ribosome slippage. His first step, then, was to develop a framework for discovering them. His technique is based on mRNA display, invented 15 years ago by Nobel laureate Jack Szostak to isolate proteins with desired functions. In this method, libraries of up to 100 trillion unique oligonucleotides (short strings of DNA) are synthesized with randomized nucleotides scattered throughout the sequence. Each DNA library member is then transcribed to RNA and tagged with an antibiotic called puromycin, which during *in vitro* protein synthesis covalently binds the RNA to the resulting protein. Using high-throughput sequencing, it then becomes possible to retroactively identify the RNA sequences that produced proteins of interest.

In order to better understand how frameshifting occurs during translation, Anzalone subjected his initial pool of oligonucleotides (268 million unique sequences) to a series of *in vitro* selection experiments. In each case he kept sequences that induced frameshifting and eliminated those that did not. mRNA display could discriminate between the two, he realized, because in order for puromycin to attach an RNA to its resulting polypeptide, the ribosome must translate to the end of the RNA. He reasoned that if a frameshift did not occur the ribosome would hit a stop codon in an mRNA. If a frameshift took place, however, the stop signal would no longer exist, making it possible for the ribosome to translate the complete RNA sequence. As a result, he could use the bound proteins as tags to enrich RNA sequences that cause a frameshift. "I found myself in a unique position," Anzalone explains, "where I was using mRNA display in a way no one else had thought of because it wasn't what it was designed for. Luckily, it happened to be very useful for what I was doing."

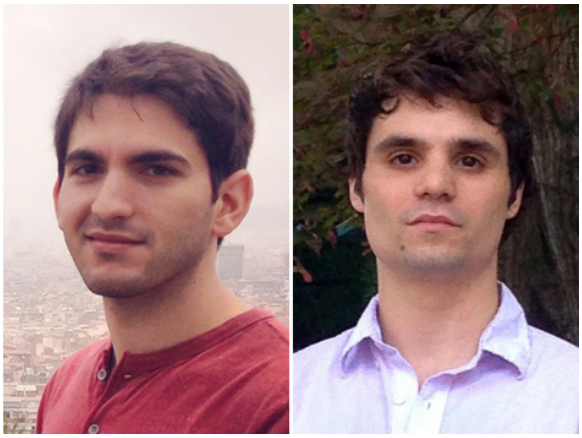
The next step was to use high-throughput sequencing to see how the RNA molecules most associated with the frameshift differed from the larger pool of sequences present before selection. Because Anzalone was trained as a chemist and not a computational biologist, however, he needed a mathematical framework for interpreting the unique dataset he had created.

Around this time he was discussing his project with Sakellarios Zairis, a friend and classmate in Columbia's MD/PhD program, and the two realized that collaborating could lead to a solution. As a member of Raul Rabadan's lab, Zairis had been working on methods for quantifying patterns of evolution in tumors from longitudinal sequencing data. Although -1 PRF exists in a completely different biological domain from cancer genetics, he recognized certain similarities in the problem. "I was already thinking about how to construct useful parameter spaces for analyzing genome evolution under strong selection," he recalls. "When Andrew first started talking about his dataset, I wondered how we could represent it in a way that would allow you to explore it intuitively."

Hypothesizing that structural features in RNA must be responsible for bumping translation out of frame, Zairis began learning about three-dimensional RNA features called pseudoknots and hairpins, in which linear sequences of nucleotides form loop-like shapes due to interactions among the nucleotides. The key insight came when he realized that the space of pseudoknots accessible to the RNAs in the selection library had seven key structural components, each of which can be summarized by an integer. Moreover, because these structures require that nucleotides on the same RNA bind to one another to keep the molecule stable, each of these parameters has a limited range of possibilities, making the space of secondary structures far more constrained and tractable than the space of primary sequences. To analyze the dataset, then, Zairis developed a mathematical pipeline capable of interpreting RNA sequencing data to identify families of pseudoknot structures that were most feasible.

Using the Department of Systems Biology's high-performance computing cluster, Zairis compared the distributions of pseudoknot geometries seen before Anzalone's selection experiments to those seen after the selection for -1 PRF. In the end, instead of needing to contend with 268 million possible pseudoknots based on sequence alone, his reduced representation condensed the dataset to approximately 2000 possible structural families. Once they identified the geometries that induced frameshifting, they went a step further, identifying the particular nucleotide preferences within those geometries, revealing instances of single substitutions that have strong effects on pseudoknot fitness.

In caffeine-fueled all night working sessions, the friends became excited by what the algorithm was unearthing, quickly gaining a clearer picture of the specific RNA motifs most capable of producing -1 PRF. Once they understood the basic biology, Anzalone could then begin incorporating these findings into molecular tools capable of producing specific frameshifts.



MD/PhD students Andrew Anzalone and Sakellarios Zairis.

Because of his interests in chemical biology, Anzalone's goal all along had been to design biochemical structures capable of changing protein translation in the presence of specific small molecules. Such structures, called riboswitches, had been designed in the past, although none had been used to induce frameshifting of the sort he had been investigating. Using a rational design approach incorporating the results of the aforementioned computational studies, Anzalone undertook a series of experiments in which he bound -1 PRF-inducing pseudoknots to small molecule-sensing aptamers. When a specific target small molecule is present, the aptamer binds to it and changes shape. This also changes the geometry of the pseudoknot, subsequently turning -1 PRF on or off. As he optimized the system, Anzalone was pleased to discover that the ratio of frameshifted proteins to non-frameshifted proteins resulting from reactions at the ribosome was very pronounced and consistent in response to the addition of a small molecule.

By combining various riboswitches, Anzalone also showed that this approach could enable a kind of biological computation, creating more complex logic gates representing AND or OR functions that could be contained on a single mRNA transcript. As a proof of principle, he designed a riboswitch for regulating apoptosis in yeast that could consistently change the ratio of viable cells to cells that undergo programmed cell death based on application of the drugs theophylline and neomycin.

Anzalone argues that riboswitches that leverage -1 PRF could offer some unique opportunities for synthetic biology. "A lot of things in biology are not about turning something on or off, but about maintaining a balance between two different regulators with opposing functions," he explains. "I realized that by using our method to set the stoichiometry and even switch between stoichiometries based on the presence of a small molecule, you could switch the phenotype of a cell." Other potential applications, he suggests, could include perturbing biological systems to understand underlying networks, incorporating labeling technologies to track individual cells, or using frameshifting motifs as sensors of specific activities inside the cell.

Anzalone's advisor, Virginia Cornish, expresses admiration at the achievement, saying, "My laboratory has been interested in engineering translation to work with unnatural amino acid building blocks, as well as in synthetic biology. So it was wonderful when Andrew came to me with the idea of bringing these two areas together. As an advisor there is nothing more satisfying that to see a graduate student achieve that level of intellectual independence."

Related publication

Anzalone AV, Lin AJ, Zairis S, Rabadan R, Cornish VW. **Reprogramming eukaryotic translation with ligand-responsive synthetic RNA switches**. Nat Methods. 2016 May;13(5):453-458.

Around the Department, 2015-2016

Selected Grants and Awards

Among the "Top 10 Papers" announced at the 2015 RECOMB/ISCB Conference on Regulatory and Systems Genomics were 4 involving Department of Systems Biology laboratories, including those of **Harmen Bussemaker**, **Richard Mann**, **Andrea Califano**, **Brent Stockwell**, and **Dana Pe'er**. **Tuuli Lappalainen** and **Michael Shen** also gave invited keynote talks at the conference.

Cory Abate-Shen was named winner of the 2015 James Family Foundation and Partner Fund Management Bladder Cancer Research Innovation Award.

Harmen Bussemaker and Coleen Murphy (Princeton) received a grant from the National Institute on Aging to dissect the genetic and molecular networks underlying longevity and aging.

Virginia Cornish received grants from the National Institutes of Health and US Department of Agriculture (with collaborators at Cornell University) to support her lab's development of a biosensor for cholera.

Arianne Giacobbe, a postdoc in the Abate-Shen Lab, was awarded the Italian Association for Cancer Research International Cancer Research fellowship and an International Cancer Alliance for Research and Education award.

Judith Kribelbauer, a PhD student in the Bussemaker and Mann labs, received a three-year HHMI International Student Research Fellowship.

Suk Hyung Lee, a postdoctoral research scientist in Michael Shen's lab, received the Urology Care Foundation 2016 Research Scholar Award.

Ankeeta Shah, an undergraduate student in Chaolin Zhang's lab, won the prestigious Barry Goldwater Scholarship and Excellence in Education Program.

Michael Shen received a three-year grant from the Department of Defense Prostate Cancer Research Program.

Peter Sims won a grant from the National Cancer Institute's Innovative Molecular Analysis Technologies (IMAT) program to support further development of single-cell sequencing efforts.

Dennis Vitkup received a grant from the NCI for "Analysis of Cancer Cell Metabolism in Diverse Environmental Conditions."

Min Zou, a postdoc in the Abate-Shen laboratory, received the 2016 Irving Institute/Clinical Trials Office (CTO) Pilot Award for "Predicting Drug Response for Human Prostate Cancer using a Cross-Species Systems Biology Approach."

New Faculty

Interdisciplinary Faculty, Department of Systems Biology
Oliver Hobert [Biochemistry and Biophysics]
Kam Leong [Biomedical Engineering]
Richard Mann [Biochemistry and Biophysics]

Center for Computational Biology and Bioinformatics
Yaniv Erlich [Computer Science]
Guy Sella [Biological Sciences]

To learn more, see **systemsbiology.columbia.edu/faculty**.

PhD Graduates

Congratulations to our recent graduates from Department of Systems Biology laboratories.

Bertrand Adanve [Cornish Lab]
Kevin Emmett [Rabadan and Wiggins Labs]
Marie Horton [Cornish Lab]
Alexander Lachmann [Califano Lab]
Jacob Levine [Dana Pe'er Lab]
Presha Rajbhandari [Califano Lab]
Forest Ray [Califano Lab]
William Shin [Califano Lab]

COLUMBIA SYSTEMS BIOLOGY

Contact Us

Columbia University Department of Systems Biology
Irving Cancer Research Center
1130 St. Nicholas Avenue
New York, NY 10032
Phone: 212-851-5208

Christopher M. Williams
Communications Director
Department of Systems Biology
cmw2189@cumc.columbia.edu

To learn more about our research and programs, visit us online at **systemsbiology.columbia.edu**.