Department of Systems Biology

Columbia University Irving Medical Center

Newsletter 2021–2022

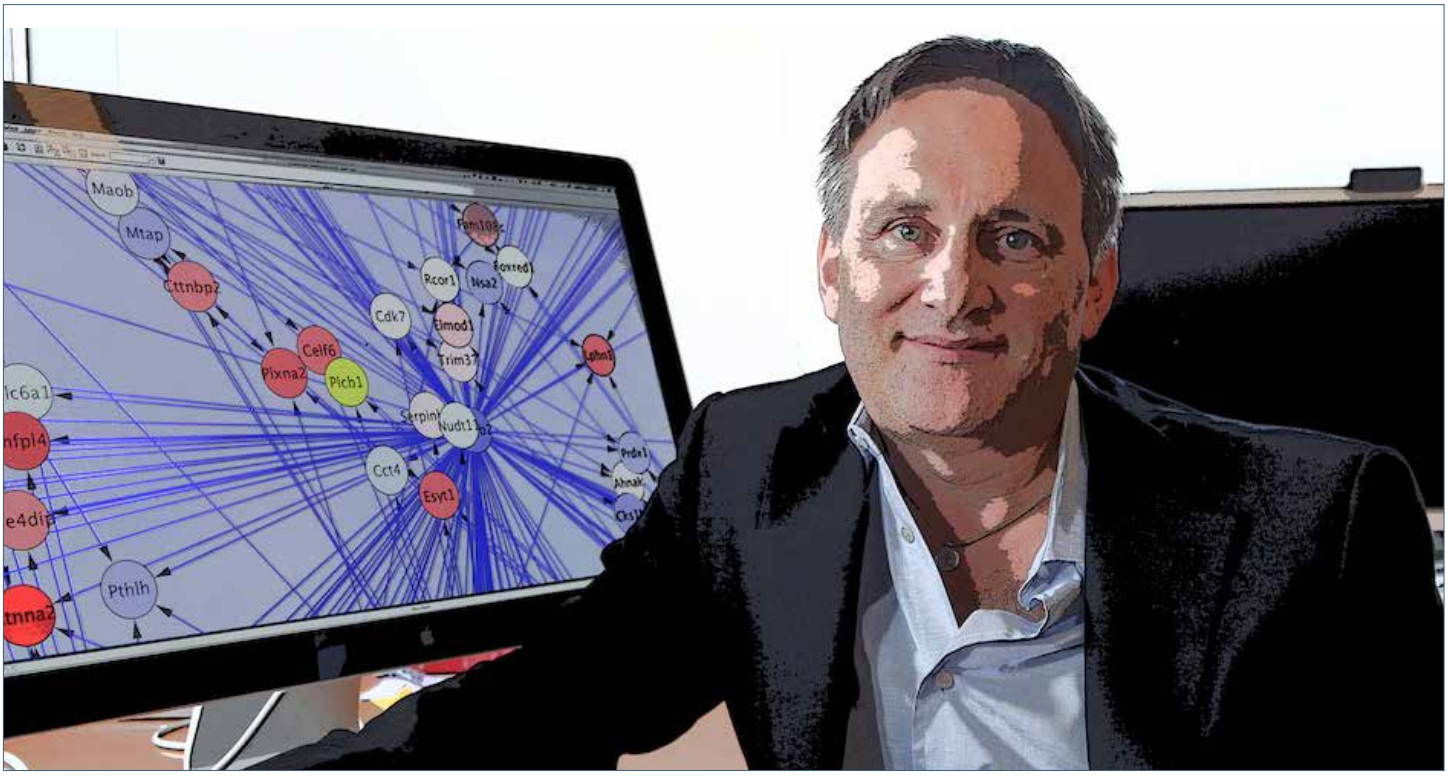
COLUMBIA UNIVERSITY
IRVING MEDICAL CENTER
Department of Systems Biology

Table of Contents

Highlights 2021–2022

Inside this issue, learn about the latest research, noteworthy grants and milestones from the Department of Systems Biology.

Deciphering Cancer Is Messy and Complex. We're Here for It.



Andrea Califano is the founding chair of the Department of Systems Biology. He directs the Columbia JP Sulzberger Genome Center and co-leads the Precision Oncology and Systems Biology program at the Herbert Irving Comprehensive Cancer Center.

Note: The following story is part of "Disruptors," a "Columbia News" series that shares innovative ideas and viewpoints from Columbia Cancer researchers and clinicians that challenge conventional thinking about cancer care, research, and beyond.

By Andrea Califano

Precision medicine has been a buzzword across the medical field for over a decade. But what does it really mean for cancer care and how is it influencing new therapies for patients? Initially, precision cancer medicine focused on targeting specific mutated genes. We thought that understanding the genetic mutations of a tumor would help us develop targeted drugs that would solve the problem, one broken gene at a time.

What we found instead is that when you build an inventory of all the broken parts in cancer, the number of mutational patterns that could give rise to cancer is larger than the number of atoms in the universe. Each mutation may even determine a different response to treatment, varying from individual to individual, and so targeting genetic mutations appears more and more as a staggeringly difficult task. Aside from the challenge created by such a vast number of possibilities, we've also discovered that genetics doesn't tell the full story of a person's cancer because cancer cells with the same exact mutations can have different drug sensitivity.

Humans have about 20,000 genes working together in ways that are different from cell to cell and from individual to individual. The enormous amount of data that we have been able to collect on cancer has helped us build computational models that, rather than trying to explain things one gene at a time, explain how all these genes work together in a system.

Move Over DNA: RNA's Role in Cancer

Approaching cancer as an ex-physicist, I wanted to open the cancer "box," look inside, and understand precisely how it works; not just one gene at a time, but based on all the gene products working together. With that, my passion has been to create the "assembly manual" of the cancer cell—a map of the complex network of molecular interactions that determine its behavior and response to treatment. This way, much like looking at the manual of a complex piece of machinery, when something is broken, we know exactly where to find the root cause and how to fix it.

The basis for this manual starts at the interface between two key molecules in the cell, DNA and RNA.

DNA, which I call the "book of what could be," contains information for all the possible things that a cell could ever be or do. In contrast, RNA represents "the book of what is" because it provides faithful copies only of the genes that a specific cell needs at a given point in time. RNA is directly translated into proteins, which are the molecules that actually "do things" in the cell, by carrying out critical functions. For instance, a liver cell and a brain cell in the same individu-

al will have the same DNA but different RNA and protein expression patterns that allow them to perform different functions.

The big paradox in cancer is that not only can the same DNA mutations produce very different RNA landscapes, but, equally important, the opposite is also true. That is, different DNA mutations can produce virtually indistinguishable RNA landscapes with identical responses to certain drugs. And the latter may hold the key to successful cancer treatment.

Yes! The problem is just as complex as it sounds.

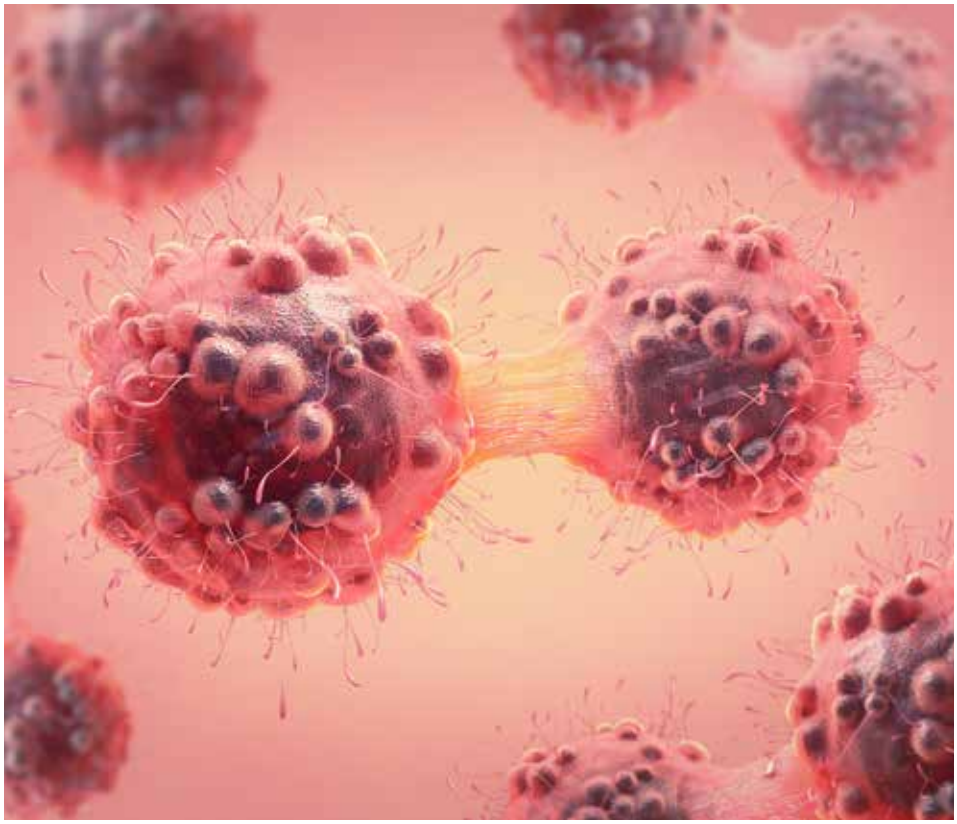
Take for example, the BRAF gene. While this is the most frequently mutated gene in melanoma, we know now that this genetic mutation can drive several other cancers, including a small subset of colon cancer, for instance. Drugs that effectively target BRAF mutant melanoma, albeit for a short amount of time, have virtually no effect in colon cancer. Same genetic mutation, yet different drug response.

Because RNA shows us specifically what is going on in a cancer cell, at a specific point in time, we have developed algorithms that accurately predict the proteins that drive the cancer cell, based only on RNA measurements. While this is far more complex than looking for DNA mutations, it also promises to be much more effective when it comes to dismantling cancer, because the protein activity state of a cancer cell provides the most informative data in terms of predicting whether a drug will kill it or not.

As Cancer Evolves So Does Our Approach to Problem Solving

In my lab, we assemble computational networks of molecular interactions between proteins and genes then analyze them to identify and target a handful of "master regulator" proteins, essentially the "pillars" that determine the cancer cell behavior and represent its most critical vulnerabilities.

We have shown that these master regulator proteins work together to power the cancer cell, akin to a building standing up on a small number of load-bearing pillars. You target one or more of these pillars and the entire building collapses. We have developed methodologies to identify precisely which proteins in each cell are



the load-bearing pillars of the cancer cell state and which drugs can best target their activities.

Precision cancer medicine has so far not fully delivered on its promise. Patients with advanced metastatic breast cancer have no more specific options than they did 10 years ago. In the next wave of precision cancer medicine, treating cancer will depend on our ability to be even more predictive with our approaches, staying one step ahead of cancer's terrifying ability to mutate or adapt to its environment, including adapting to therapies.

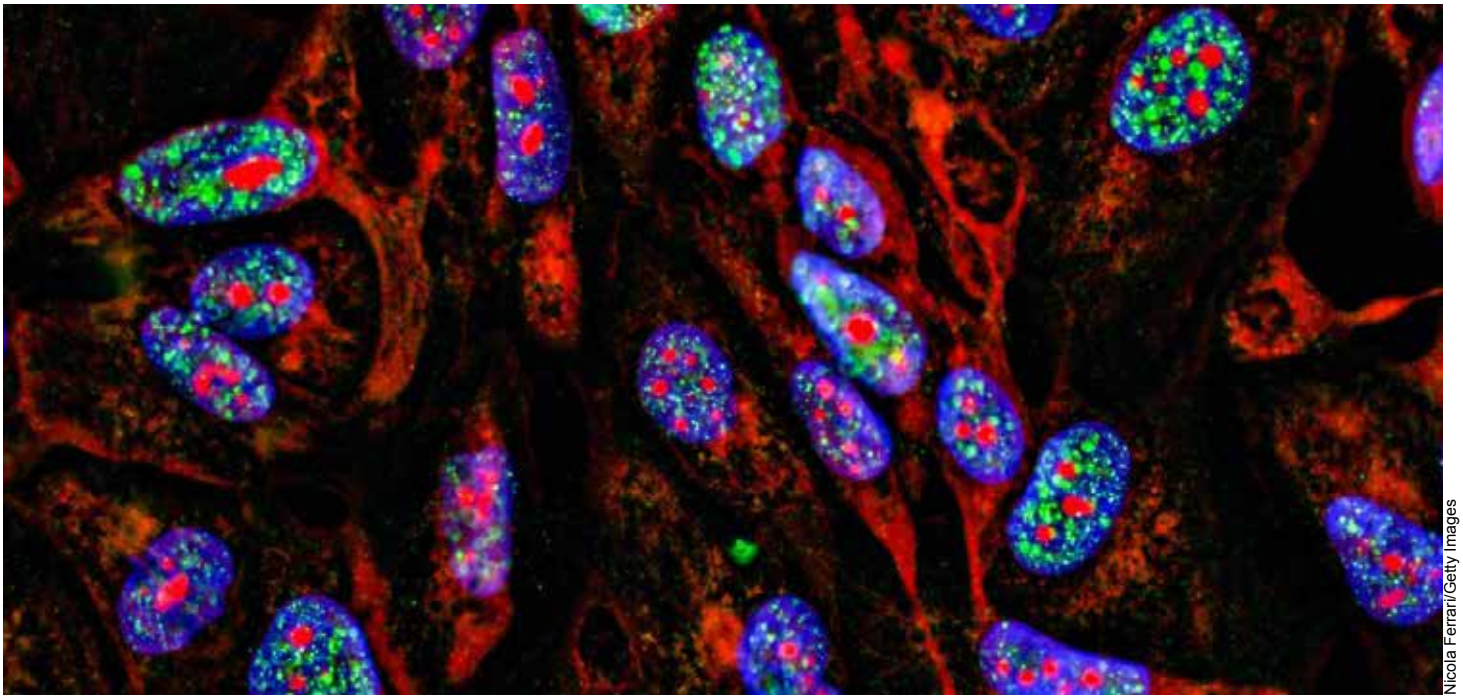
Beating Cancer at Its Game to Mutate and Thwart Treatment

The only way to stay one step ahead is to surrender to the complexity of cancer and understand that each cancer cell is different, even the same cancer in the same patient. Our algorithms identify groups or types of cancer cells that will respond to certain treatments so that we will know which cells are killed and which are spared when we treat with a specific drug. In that way we can identify the combination and sequence of drugs that will kill each of those specific population of cells within an individual patient's tumor, removing the guess work to identify drugs or drug combinations that are truly personalized to the patient. This moves us past treating cancers one genetic mutation at a time to treating it as complex networks of broken genes and proteins that determine their drug response.

Cancer is complicated and cannot be simplified beyond a certain level in our research, just as we cannot simplify cancer for our patients. We need to embrace its complexity, matching its intricacy and sophistication with approaches that are equally complex and sophisticated.

Andrea Califano is the founding chair of the Department of Systems Biology, the Clyde and Helen Wu Professor of Chemical and Systems Biology, professor of biomedical informatics and biochemistry, and molecular biophysics, and professor of medicine at Columbia University Vagelos College of Physicians and Surgeons. He directs the JP Sulzberger Columbia Genome Center and co-leads the Precision Oncology and Systems Biology program at the Herbert Irving Comprehensive Cancer Center.

Study Shows Why Many Cancer Cells Need to Import Fat



Nicola Ferrari/Getty Images

Columbia and MIT researchers are revealing the surprising reasons why cancer cells are often forced to rely on fat imports, a finding that could lead to new ways to understand and slow down tumor growth.

The research, led by Dennis Vitkup, PhD, associate professor of systems biology at Columbia University Vagelos College of Physicians, and Matthew Vander Heiden, MD, PhD, director of the Koch Center at MIT, was published June 23 in Nature Metabolism.

Common nutrients we eat, like fat, and the oxygen we breathe, are likely to play an essential role in the growth of cancer cells.

Oxygen is most known for its role in making energy in the body; that is why when we exercise, we start breathing harder. Because many cancer cells live in oxygen-depleted environments, it is often assumed that their growth is limited by energy.

But oxygen also has a less-celebrated role, and that is to provide oxidizing power for the chemical reactions driving synthesis of biomolecules necessary for building new cells. Many biosynthetic reactions require a co-factor called NAD+. When oxygen is lacking, cells cannot regenerate growth-promoting NAD+, and their key synthetic reactions come to a halt.

The new study found, surprisingly, that hypoxic cancer cells usually have more energy than they need for growth. When the researchers provided cancer cells with extra nutrients for energy generation, the cells did not respond.

Instead, when researchers used various methods to unclog biosynthetic pathways inhibited by lack of oxygen, cancer cells robustly increased proliferation.

The researchers found that while various biosynthetic pathways are sensitive to oxygen availability, synthesis of fats was among the most affected. Fat molecules are used to create membranes of new cells, and fat synthesis is especially challenging for cancer cells that need to synthesize new membranes for their growth. Without access to oxygen, cells cannot adequately supply their fat synthesis pathways.

"What makes our result very counterintuitive," Vitkup says, "is that fat synthesis is not considered to be a process requiring a lot of oxygen. But our experiments demonstrated that up to 30% of oxygen used by cancer cells is not for energy generation but for synthesizing fats."

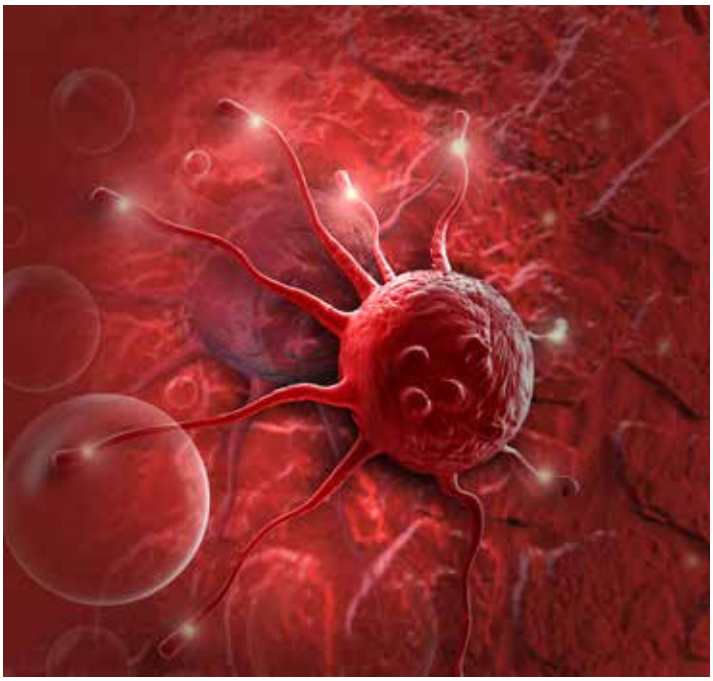
As a result of oxygen's impact on biosynthesis, cancer cells growing in oxygen-limited environments are strongly dependent on the import of fats from the environment. This creates a crucial vulnerability for cancer cells, such that cutting their supply of imported fats may slow or stop cancer growth.

Vitkup's team is now trying to identify the receptors that cancer cells use to import fats in different tumors and which receptors could be targeted by drugs. The study also suggests that changing the composition of fats in the diet may play a vital role in influencing cancer growth.

"We usually think of cancer as being driven primarily by genetic mutations, but for cancer cells living in challenging conditions, such as oxygen-starvation, their environment is equally important," Vitkup says. "Mutations stimulating uptake of fats, for example, will only promote tumor growth if these fats are actually available in their environment."

Precision Medicine Sends Cancer for a Loop

By Alan Dove



In a tour de force of precision medicine, an international team led by two Columbia University Herbert Irving Comprehensive Cancer Center (HICCC) researchers has demonstrated the potential of a new approach to prostate cancer treatment. The iterative strategy, named OncoLoop, uses a sophisticated computer algorithm to match each patient to an "avatar"—a carefully engineered laboratory model of a cancer that matches the molecular profile of the patient's tumor. By selecting drugs that target the master regulator proteins responsible for the genetic program driving the avatar's tumor, the team can then validate the treatment regimen most likely to work in that patient.

"We take a patient, we match them to the set of mouse tumors we already have, and to the set of drug data that we already have, and then actually identify drugs that are most likely to work in both the patient and the model … test those drugs in those models, and then use that information to inform clinical treatment," says Cory Abate-Shen, PhD, a senior author on the new work and member of the HICCC and department chair of Molecular Pharmacology and Therapeutics at Columbia University Vagelos College of Physicians and Surgeons (VP&S).

The strategy solves a fundamental problem in cancer drug research: cellular and animal models of the disease are often considered to be poor mimics of any particular patient's tumor. "There's been tremendous controversy around what is a good model, and so we decided that maybe we needed a more quantitative and sophisticated approach," says Andrea Califano, Dr, chair of the Department of Systems Biology at VP&S and co-leader of the Precision Oncology and Systems Biology program at the HICCC. Califano is also a senior author on the new paper, published in *Cancer Discovery.*

To accomplish that, Abate-Shen, Califano, and their collaborators at Columbia University Irving Medical Center and several other research institutes, combined algorithmic mining of large sets of patient data with a collection of genetically engineered mice that replicate different types of genetic dysregulation found in prostate cancers, providing researchers with dozens of models of prostate cancer that can more closely mimic the heterogeneous tumor types found in patients. Until recently, though, that would have seemed like an impossible task.

"If you look at the number of genes whose mutations are potentially associated with cancer, there are about 2,000 of them; if you look at all the possible ways in which you could combine them, you get more potentially tumorigenic patterns than there are atoms in the universe," says Califano. However, the researchers have shown that the majority of such tumorigenic patterns are processed by a limited set of master regulator proteins.

"Think of cancer mutations as grains of sand in the upper part of an hourglass," says Califano, adding that "you don't need to shoot every single grain of sand, because, if you just block the middle point they all have to go through to reach the bottom, the hourglass will stop working."

In the OncoLoop approach, the investigators analyze the gene regulation profile of a patient's cancer to determine which master regulators are involved, and select an avatar model with the same master regulator profile. They can then identify drugs likely to invert the activities of those master regulators to kill the tumor cells instead of preserving them, and test those drugs individually or in combination on the patient's avatar model to validate them.

Besides providing treatment plans that attack the patient's tumor most effectively, the strategy makes it less likely that the cancer will evolve resistance to the regimen. "By getting drugs that invert the master regulator program, you're not hitting on one particular target, [but] the focal point, and it becomes very, very hard for the cell to get around that," says Abate-Shen. The limited number of master regulators also means that once the team has identified an effective treatment for a given profile, it should work in other patients who share the same profile.

While prostate cancer provided an ideal system in which to prove that the concept works, OncoLoop could see broader use. "We are already running a number of clinical trials where we're using these [methods] in other tumors," says Califano.

Both researchers emphasize that cross-disciplinary collaboration was crucial to the project's success. "Our models are great models, but they wouldn't be as valuable as they are without the algorithms, and the algorithms are really valuable because they're applied to biologically relevant models," says Abate-Shen. Califano adds that "that's why we've been working together for the past twelve years."

Chaolin Zhang Focuses on RNA Regulatory Code and RNA-Based Precision Medicine

As an engineer converted to biologist, Chaolin Zhang, PhD, was drawn to understanding gene regulatory logic. Early in his academic career, he decided to tackle the RNA code, i.e., how specific interactions between RNA-binding proteins (RBPs) and their target transcripts determine gene output, ultimately manifested in dynamic systems in health and disease conditions.

Zhang is an associate professor of systems biology and biochemistry and molecular biophysics. After training as an automation engineer with a focus on machine learning at Tsinghua University, in Beijing, he completed his PhD work in computational biology at Cold Spring Harbor Lab, followed by postdoctoral training at Rockefeller University. He joined Columbia as an assistant professor in 2012 and received tenure in 2019.

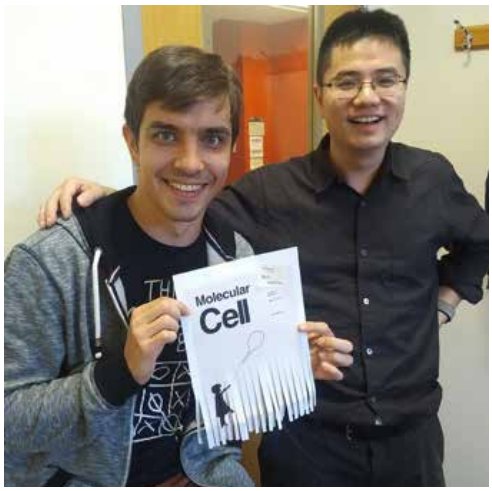
His team regularly applies a variety of tightly integrated experimental and computational approaches and techniques to both cell-based and mouse models. These include CRISPR-based genome engineering, high-throughput screening, deep sequencing, probabilistic modeling, and machine learning. The lab's dry lab and wet lab are next to each other, to facilitate a multidisciplinary approach to tackling research questions.

When starting his lab at Columbia, Zhang decided to focus on the regulation of alternative splicing, a molecular process that removes introns and rejoins exons in precursor mRNA in the brain during neurodevelopment. His lab identified thousands of alternative exons, showing dramatic switches at precise times during development, as well as major RBP splicing factors that are turned on to control such developmental switches. Since the brain is composed of hundreds of different cell types, including many subtypes of neurons and glial cells, the group recently took advantage of massive single-cell RNA-seq data to delineate specific regulation in individual cell types.

"Using models that differentiate pluripotent stem cells into neurons in a petri dish," says Zhang, "allowed us to manipulate individual regulators and even single alternative exons. This enabled us to determine key regulatory events that are instrumental to fundamental properties of nerve cells, such as how neurons get excited."

"Now," adds Zhang, "we want to view RNA regulation from an evolutionary perspective, as many alternative splicing events diverge in different species, including humans and other primates, during evolution. Such genomic changes are double-edged swords. While they may underlie our unparalleled cognitive capabilities, they also render us more susceptible to a variety of neurodevelopmental and neurodegenerative diseases."

Let-7 microRNA partially escapes recognition and suppression by LIN28, a finding based on a new RNA-binding mode of LIN28 discovered by the Zhang lab. This *Molecular Cell* cover art, inspired by Bansky's mural "Girl with Balloon" (2002), was designed by Dmytro Ustianenko, a former postdoc in Zhang's lab and first author on the journal paper.



At a live Sotheby's auction in 2018, a painting of "Girl with Balloon" sold for $1.4 million. A device hidden in the painting's frame then activated, partially shredding the work. In response, a Danish professor sent a shredded journal cover to Zhang's lab as a gift. From left: Dmytro Ustianenko and Chaolin Zhang.

In parallel with its efforts to understand the functional significance of RNA regulation, the Zhang lab is also probing the mechanistic side—how RBPs recognize their substrate transcripts. Protein-RNA interactions are building blocks of the RNA code. Despite exciting progress in mapping protein-RNA interactions with cross-linking immunoprecipitation (CLIP) and other related techniques, scientists still have a poor understanding of RBP specificity. Zhang's group has harnessed the "signature" left behind by crosslinking protein and RNA by UV, so they can map protein-RNA interactions at single-nucleotide resolution.

"We also developed statistical models," says Zhang, "to take full advantage of these high-resolution maps to better define the binding specificity of RBPs. We were able to reveal novel binding modes for RBPs that had been studied for decades, with implications for stem cell biology and fundamental mechanisms of alternative splicing regulation."

Moving forward, the Zhang lab plans to use novel experimental and AI-based computational approaches to investigate such aspects of protein-RNA interactions as protein-RNA complex structures, splicing regulatory elements embedded deep in introns, and the impact of genetic mutations on protein-RNA interactions.

Though a major focus of his current research is fundamental biology, Zhang had the opportunity to witness the transformative impact of RNA-based precision medicine while working at Cold Spring Harbor Lab with Dr. Adrian Krainer, who developed the first splicing modulating antisense oligo (ASO) drug SPINRAZA (nusinersen), to treat spinal muscular atrophy. This was a

Continued from page 7

perfect example of how a fundamental understanding of RNA regulation can be translated into life-saving drugs.

About 10 percent of the human population are affected by single-gene diseases, including many neurodevelopmental disorders, most of which currently have no treatment. Zhang's group has been developing technologies to identify drug targets for such unmet medical needs and to speed up the ASO drug-screening process.

"Modulating alternative splicing," says Zhang, "provides a molecular mechanism not only for diversifying protein products, but for eliminating erroneous splicing variants or restoring the deficient genes for therapeutics. I am very excited about this new goal."

"The COVID-19 pandemic and development of mRNA vaccines have brought RNA research into the spotlight," he adds. "The technologies have moved forward very dramatically in recent years. This is an opportune time to advance our understanding of RNA biology and transform that understanding to life-saving therapeutics."

REFERENCES

Feng H, Moakley D, Chen S, McKenzie, MG, Menon V, Zhang C. Complexity and graded regulation of neuronal cell type-specific alternative splicing revealed by single-cell RNA sequencing. Proc Nat Acad Sci 2021 Mar 09;118(10). doi:10.1073/pnas.2013056118.

Jacko M, Weyn-Vanhentenryck SM, Smerdon JW, Yan R., Feng H, Williams DJ, Pai J, Xu K, Wichterle H, Zhang C. 2018. Rbfox splicing factors promote neuronal maturation and axon initial segment assembly. Neuron 2018 Feb 21;97(4):853–868.e6. doi: 10.1016/j.neuron.2018.01.020.

Ustianenko D. Chiu HS, Treiber T, Weyn-Vanhentenryck SM, Treiber N, Meister G. Sumazin P, Zhang C. LIN28 selectively modulates a subclass of let-7 microRNAs. Mol. Cell. 2018 Jul 19: 71(2): 271–283.e5. doi: 10.1016/j.molcel.2018.06.029.

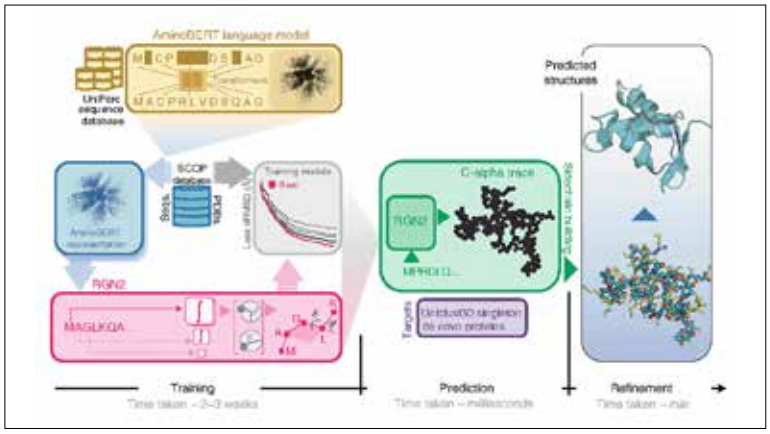
Weyn-Vanhentenryck SM, Feng H, Ustianenko D, Duffié R, Yan Q, Jacko M, Martinez,JC, Goodwin M, Zhang X, Hengst U, Lomvardas S, Swanson MS, Zhang C. Precise temporal regulation of alternative splicing during neural development. Nat Commun 2018 June 06;9:2189.

Single-Sequence Protein Structure Prediction Using a Language Model and Deep Learning

For a long time, work in predicting 3D protein structures from amino acid sequences has relied on physics-based methods that estimate energy landscapes and dynamically fold proteins within these landscapes. Then about ten years ago, researchers shifted to using information from co-evolutionary relationships embedded in multiple sequence alignments (MSAs). More recently, AlphaFold2 and similar AI systems have been hailed for predicting a protein's structure using deep learning and co-evolutionary relationships encoded in MSAs. These systems are less successful, however, at predicting the structure of orphan proteins, which have few if any homologs, or of rapidly evolving proteins for which an MSA cannot be generated. Nor are they good at understanding the rules governing spontaneous polypeptide folding in solution.

Now Mohammed AlQuraishi, assistant professor of systems biology, and his colleagues have taken a different path entirely and developed a recurrent geometric network (RGN) algorithm, called RGN2, which uses a deep-learning language model (AminoBERT) that avoids the use of alignments completely. RGN2 shows that language models are effective at learning structural information from primary sequences while being able to extrapolate beyond known proteins, enabling them to predict orphan and designed proteins.

RGN2 outperforms AlphaFold2 and RoseTTAFold on predicting the structure of orphan proteins, as well as designed proteins, which by definition have no homologs. A linked geometric module compactly represents Cα backbone geometry in a translationally and rotationally invariant way. In addition, their algorithm requires a 106-fold reduction in computer time. An added benefit of using a deep-learning language model is the potential for generating new functional proteins.



Organization and application of RGN2. RGN2 combines a transformer-based protein language model (AminoBERT, yellow) with an RGN that uses Frenet–Serret frames to generate the backbone structure of a protein (green). After initial construction of the sidechains and hydrogen-bonded networks, refinement of the structure is subsequently performed using AF2Rank (blue).

An algorithm that predicts structure directly from a single sequence is closer to the real physical process than one that uses MSAs and has the potential to lead to new understanding of protein biophysics. Rapid predictions for large numbers of long proteins would enable many practical applications in enzymology, therapeutics, and chemical engineering.

Continued from page 8

"Many systems biology models," says AlQuraishi, "tend to abstract away details of the underlying molecules and the interactions that underpin systems-level phenomena. This makes such models 'brittle' and unable to generalize well, as they don't leverage the underlying commonality of all molecules—that they're made of atoms and present surfaces that define their interactions with other molecules.

"RGN2 and methods like it help bridge the knowledge gap between protein sequence and structure by enabling prediction of many protein structures at scale. This, in turn, will enable development of future systems biology models that are explicitly structural in nature, i.e., that don't abstract away the molecular details."

Language models will not replace MSAs, but offer a complementary way to determine the latent rules that govern protein folding. AlQuraishi foresees the development of methods able to compute sequence-to-structure maps without requiring explicit evolutionary information.

The work was featured as the cover story of the November 2022 issue of *Nature Biotechnology*.

REFERENCE

Chowdhury R, Bouatta N, Biswas S, Floristean C, Kharkar A, Roy K, Rochereau C, Ahdritz G, Zhang J, Church GM, Sorger PK, and AlQuraishi M. Single-Sequence Protein Structure Prediction Using a Language Model and Deep Learning. *Nature Biotechnology* 2022 Oct. 03. doi.org/10.1038/s41587-022-01432-w.

New Machine Learning Method Predicts Damaging Missense Variants

The premise of genomic medicine is that a person's genomic characterization can be used to improve medical diagnosis, prognosis, and treatment. Each person, however, has millions of genetic variants, the vast majority of which have negligible impact on their health. How to determine which variants are relevant to a particular condition is a central issue in genomic medicine.

The issue is most pressing in the case of missense variants, which alter a single amino acid in proteins. Only about 20–30 percent of these mutations have a functional impact. Thus the question of how likely a variant is to change protein function—contributing to a health condition—is extremely uncertain for missense variants. As a result, most missense variants in clinical genetic testing are classified as VUS (variant of uncertain significance).

Yufeng Shen, PhD, an associate professor in the Department of Systems Biology and the Department of Biomedical Informatics, and his group have developed a new method for predicting which missense variants are potentially damaging. The method, called gMVP (graphical model for predicting Missense Variant Pathogenicity), uses one of the latest machine learning techniques, a graph attention model, to capture information relevant to predicting which variants are potentially damaging. Their paper, "Predicting Functional Effect of Missense Variants Using Graph Attention Neural Networks," was published in *Nature Machine Intelligence* on November 15, 2022.

The new method uses the coevolution of pairs of amino acid positions in a protein to determine whether a pair is functionally correlated. This makes it possible to pool information across functionally correlated positions that are not close in sequence. Shen and his group used several independent data sets representing different applications of the method, including clinical genetic tests and new disease gene discovery, to evaluate the performance of the method. In all the tests, gMVP substantially out-performed other methods.

"Predicting the effects of missense variants," says Shen, "has been well studied. We usually avoid well-studied problems. But this problem is far from being solved and, more important, even small improvements can lead to real changes in genomic medicine. Anytime we can confidently change the interpretation of a VUS to a damaging variant for a patient, it has the potential to lead to more effective treatment."

"We've also had a lot of fun," adds Shen, "learning the latest machine learning techniques and trying out different modeling methods."

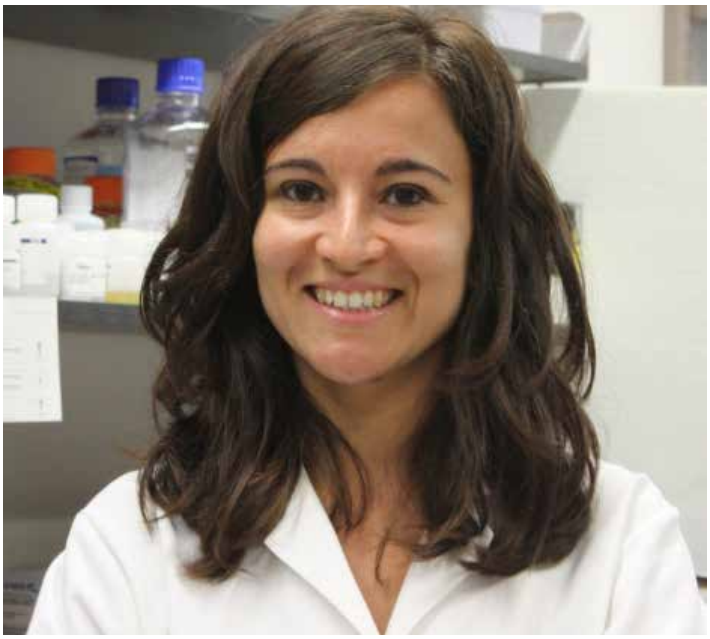
REFERENCE

Zhang H, Xu MS, Fan X, Chung WK, Shen Y. Predicting functional effect of missense variants using graph attention neural networks. *Nat Mach Intell* 2022 Nov 15;4:1017–1028. doi.org/10.1038/s42256-022-00561-w.

FACULTY SPOTLIGHT

Q&A with Sara Zaccara

Sara Zaccara is an assistant professor in the Department of Systems Biology. She grew up in a small town in Italy and received her bachelor's and master's degrees in biotechnology from the University of Florence. She received her PhD from the University of Trento, in Trento, Italy, where her thesis work was on p53-dependent translational regulation. In fall 2022, Zaccara came to Columbia after a postdoctoral position at Weill Cornell Medicine.



Q: *Please describe your interest in mRNA degradation in normal and disease states.*

A: We know that messenger RNA (mRNA) is the output of a well-controlled and regulated transcriptional program. Until not too long ago, it was thought that cells adapt their transcriptional program to changes in their environment while simply using RNA degradation as a mechanism to discard defective molecules resulting from mistakes in transcription or to discard molecules no longer needed. However, this model turns out not to be accurate. RNA degradation controls the overall cellular balance as much as transcription does. We now have more and more evidence that cells continuously adapt to environmental changes in normal and disease states by modulating their mRNA degradation program.

Over the years my work—and that of others—has extensively characterized numerous mechanisms, such as the epitranscriptome, that are involved in regulating RNA degradation. More and more studies highlight the heterogeneity of the mechanisms involved in this process. This complexity has limited our ability to disentangle the precise mechanisms involved in mRNA dysregulation, particularly during the onset of complex multifactorial events, such as many diseases. The goal of my lab is to disentangle this complexity so as to understand the fundamental rules governing the dynamics of mRNA degradation.

Q: *What exactly is the epitranscriptome?*

A: The "epitranscriptome" refers to a newly elucidated post-transcriptional mechanism regulating mRNA degradation. It implies the reversible and dynamic chemical modification of mRNA, analogous to phosphorylation in proteins. More than any other nucleotide modification, that of adenosine methylation to form N6-methyladenosine (m6A) has captured the attention of the scientific community. This is because (1) m6A is the most abundant modification in the mRNA transcriptome and (2) compared with other modifications, it has emerged as a widespread regulatory mechanism that controls the transcriptome to mediate diverse crucial physiological processes.

Q: *How does m6A contribute to mRNA degradation?*

M6A is a clear mark for mRNA degradation. My recent work has been instrumental in clarifying the mechanism by which m6A mRNA degradation occurs in cells. We now know that YTHDF proteins, which bind the m6A-modified mRNAs, are the key players to mediate m6A mRNA degradation. We also found that all YTHDF proteins contribute to m6A biology in the same way, ensuring that degradation takes place. We have learned that we can target all of them and don't need to find drugs for each of them. Right now, the m6A writer, METTL3, is the target of therapies. More and more companies are focusing on it. Based on our work, YTHDF proteins could be a target, as well.

Q: *What do we still not know about m6A mRNA degradation?*

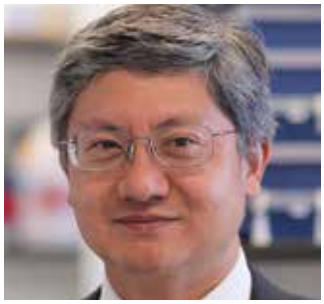
A: We still need to understand how the m6A mRNA degradation process is regulated and how it plays a critical role in the overall RNA degradation programs used by the cells in different cellular conditions. To answer these questions, we plan to take unbiased multidisciplinary approaches. We are excited at the idea of using CRISPR as a tool to precisely edit YTHDF readers. This will enable us to interrogate these proteins at single-amino acid resolution and ask questions such as, "Which amino acids are functional sites, and how do they contribute to regulating RNA degradation?"

On a different note, we should keep in mind that RNA is becoming the molecule of the future. We all know that RNA is the key molecule of our recent RNA vaccine. But not everyone knows that this RNA molecule has modifications that are instrumental to ensuring that it activates the correct response in our body and stays in our body for the amount of time needed. In other words, it needs to be degraded at the right time.

The RNA modification field started only 10 years ago. So there is more and more to explore about how these modifications could be used for future RNA vaccines. To do this, we need to answer basic questions about how RNA modifications are regulated in cells and about the interplay between these modifications and other RNA regulatory mechanisms. This is what I have worked on so far, and it is what I want to do in the future: characterize these RNA degradation mechanisms in the hope that they can help us to develop even more ways to use them for future forms of therapy.

NEWS

Molly Przeworski, PhD, Biological Sciences and Systems Biology, was appointed by the National Academies of Sciences, Engineering, and Medicine to serve on the Committee on the Use of Race, Ethnicity, and Ancestry as Population Descriptors in Genomics Research.



Michael Shen, PhD, the Arthur J. Antenucci Professor of Medical Sciences and professor of genetics, urology, and systems biology at Columbia's Vagelos College of Physicians and Surgeons, has been named a fellow of the American Association for the Advancement of Science (AAAS). He is being recognized for his contributions to developmental and cancer biology, particularly his work in applying developmental genetics to the study of genitourinary cancers.

Annual Retreat

The Department of Systems Biology's 2022 annual retreat was held September 29–30 at Honor's Haven, in Ellenville, NY. It was the first to be held post-COVID; the last one was held in 2019. The three poster winners were Amirhossein Jafariyan, for "Live Yeast Dressings Engineered via Synthetic Biology to Deliver Protein Factors to Diabetic Wounds"; Zhouzerui Liu, for "SCOPE-seq Analysis of Intra-operative Labeling Specificity for Glioblastoma"; and Jiayao Wang, for "Mutations Associated with Autism Spectrum Disorder Affect Cohesive and Functionally Interrelated Brain Circuits."

Top: Honor's Haven grounds.
Bottom, from left: poster presentation by Lorenzo Tomassoni (Califano lab); winner of poster presentation, Zhouzerui Liu (Sims lab); Sara Zaccara.

Around the Department, 2021-2022

Grants, Awards, and More

Benjamin Izar, MD, PhD, joined the Systems Biology faculty.

Itsik Pe'er, PhD, was promoted to the rank of professor.

Sara Zaccara, PhD, joined the Systems Biology faculty.

Corrine Abate-Shen, PhD, Molecular Pharmacology and Therapeutics, will receive $486,000 from the Department of the Army, Army Medical Research and Material Command, for "Role of ATAD2 in Prostate Cancer Bone Metastasis."

Andrea Califano, Dr, will receive $6,909,000 over seven years from the National Cancer Institute for "Predicting Cancer Cell Response to Endogenous and Exogenous Perturbations at the Single Cell Level."

Andrea Califano, Dr, will receive $4,893,902 over five years from the National Cancer Institute for "Elucidating and Targeting Tumor Dependencies and Drug Resistance Determinants at the Single Cell Level."

Aleksandar Z. Obradovic received the 2022 Dean's Award for Excellence in Research, in recognition of the high quality and significance of his PhD thesis research in the Coordinated Doctoral Programs in Biomedical Sciences.

Raul Rabadan, PhD, will receive $6,804,000 over seven years from the National Cancer Institute for "Towards a Quantitative Understanding of Tumor Evolution" and $649,998 over two years from STAND UP TO CANCER for a subaward of "Multiomic Analysis of Immune System and Microbiota Influence on Temporal and Spatial Evolution of Tumor Microenvironments."

Four Columbia research teams have been awarded the inaugural Edward P. Evans Center for Myelodysplastic Syndromes (MDS) Pilot Awards and Fellowships. Each team will receive a one-year $100,000 grant for the Edward P. Evans Center Pilot Awards and a two-year $60,000/year grant for the fellowships. The two pilot projects are being led by principal investigators **Pawel Muranski,** MD, assistant professor of medicine and of pathology and cell biology at Columbia University Vagelos College of Physicians and Surgeons (VP&S); **Amer Assal,** MD, assistant professor of medicine at VP&S; and **Raul Rabadan**, PhD, professor of systems biology and of biomedical informatics at VP&S. Rabadan's project, with co-investigator **James Manley**, PhD, professor of life sciences, is "Systematic Protein Structure Characterization of Mis-spliced Transcripts in Myelodysplastic Syndromes."

Rodney Joel Rothstein, PhD, Genetics and Development, will receive $3,924,540 over five years from the National Institute of General Medical Sciences for "Molecular Mechanisms Underlying Recombination at DNA Double-Strand Breaks and Stalled Replication Forks."

Michael Shen, PhD, Genetics and Development, will receive $1,950,515 over five years from the National Institute of General Medical Sciences for "Genetic Approaches to Development and Disease."

Peter Sims, PhD, and Donna Farber, PhD, will receive $2,275,349 over three years from the Wellcome Sanger Institute for a subaward of "Single Cell Analysis of Aging in Lymphoid Compartments."

Harris Wang, PhD, and Samuel Sternberg, PhD, will receive $2,665,170 over four years from the National Institute of Biomedical Imaging and Bioengineering for "A High-Performance and Versatile Technology for Precision Microbiome Engineering."

Xuebing Wu, PhD, Medicine, will receive $600,000 over three years from the Pershing Square Sohn Cancer Research Alliance for "Targeting Tumor with CRISPR/Cas13-Based Sequence-Specific Cell Knockout."

Chaolin Zhang, PhD, will receive $1,200,000 over three years from the Simons Foundation Autism Research Initiative via the 2021 Genomics of ASD: Pathways to Genetic Therapies Initiative.

Chaolin Zhang, PhD, will receive $3,006,392 over five years from the National Institute of Neurological Disorders and Stroke for "RNA Regulatory Networks in Neuronal Cell Type Diversity and Function."

Chaolin Zhang, PhD, will receive $568,846 from NIH/NHGRI for "Mapping Proximal and Distal Splicing-Regulatory eEements."

Chaolin Zhang, PhD, will receive $4,313,684 over five years from NIH/NIGMS for "Complexity and Evolution of Splicing-Regulatory Networks." He will also receive $2,066,788 over four years from the NIH/NIGMS for "Integrative Analysis of Tissue-Specific Alternative Splicing Regulation Under Adaptive Selection."

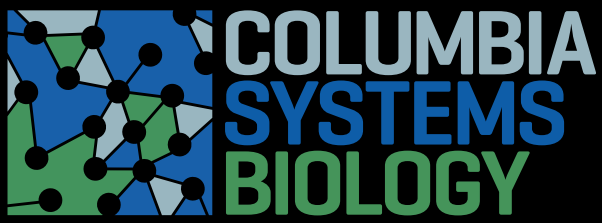
Chaolin Zhang, PhD, will receive a Scientific Innovation Award of $150,000 over two years from the Brain Research Foundation.

PHD GRADUATES

Congratulations to our recent grads!

Nicholas Giangreco (Tatonetti lab)
Benjamin Hobson (Sims lab)
Yiming Huang (Wang lab)
Jordan Kesner (Wu lab)
Jordan Metz (Sims lab)
Aleksandar Obradovic (Califano lab and Drake lab)
Felix Wu (Przeworski lab)

COLUMBIA UNIVERSITY
IRVING MEDICAL CENTER



Columbia University Department of Systems Biology
Irving Cancer Research Center
1130 St. Nicholas Avenue
New York, NY 10032

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

Design by: Nicoletta Barolini
Edited by: Ann Rae Jonas