THE
DARK GENOME'S EXPLORER

Howard Chang's discovery of the body's cellular GPS drew him into the vast, dark expanse of the noncoding genome, exploring its control of gene expression and how its dysfunctions fuel multiple cancers.

In 2001, Howard Chang was planning a series of experiments to examine how aging alters gene expression in skin cells known as fibroblasts when he noticed that an important control was missing. Skin scientists, it seemed, had implicitly assumed that one fibroblast—which churns out fibrous proteins and other macromolecules that build the scaffolding of tissues—is pretty much like any other. Chang, who had recently started a postdoctoral fellowship in Patrick Brown's laboratory at Stanford University, realized this assumption needed checking: If he was going to compare gene expression in cultured fibroblasts with that of their counterparts from more aged and different anatomical sites, he needed to know how similar they were to begin with. "It turned out," Chang recalls, "that fibroblasts from different parts of body were as different from each other as different types of white blood cells."

In subsequent studies conducted in Brown's and, later, his own lab, Chang discovered a key cause of those differences: the humble fibroblast, it turned out, is a vital component of the body's global positioning system. "Fibroblasts have positional memory and gene expression programs that are distinct based on where they come from in the body," says Chang, who has since 2017 been the Virginia and D.K. Ludwig Professor of Cancer Genomics at Stanford. "They retain that information and then share it through signaling to other surrounding cell types." In exploring the genomic source of this anatomical GPS, Chang would wander into the darkest regions of the genome—that 98% of the whole that encodes no proteins but controls which genes in the remaining 2% are expressed. Along the way, he and his team devised powerful new technologies to probe the dark genome, detailing how it controls programs of gene expression in health and disease.

HOWARD CHANG

LUDWIG STANFORD

> **"If we're saying that different parts of the skin have beautifully laid out address codes, a cancer cell going from one part of the body to another clearly has to deal with those address codes."**

In 2018, Chang and his Stanford colleague William Greenleaf reported in *Nature Medicine* their use of the latest of those technologies, ATAC-Seq, to characterize the subtly different states assumed by seemingly identical T cells of the immune system. The technique, they showed, could also be used as an important diagnostic tool in treating a T cell leukemia that manifests in the skin. But that was just for starters. Later in the year, in partnership with researchers from multiple institutions, they reported in *Science* a granular map of the dark genome's regions that are open for business in 23 types of tumors, revealing how mutations alter the landscape of the genome and patterns of gene expression in each to activate cancer-driving genes, and why subtle variations in noncoding DNA sequences predispose people to various cancers.

Apprenticeships

Chang was born in Taipei, Taiwan, and emigrated to the U.S. when he was 12 years old. Though his was always an academic-minded home—his father was a physician—Chang says he became interested in science only after arriving in the U.S., when he

and his friends proposed projects for the science fair that required instruments not typically found in a high school laboratory. His biology teacher introduced him to a friend at the University of California, Irvine, whose lab focused on transplantation biology. It was there that Chang got his first taste of scientific research.

After his freshman year at Harvard University, Chang did a summer research stint in the laboratory of the enzymologist and former Ludwig scientific advisor Christopher Walsh exploring the mechanism of action of the transplant rejection drug cyclosporin A. "Over the course of that summer, I realized that what people had thought about how this drug works was about to be transformed," Chang recalls. "That was tremendously exciting. It was one of the reasons I became interested in fundamental research." The interest solidified into a plan after he spent two formative summers in the Undergraduate Research Program at Cold Spring Harbor Laboratory, where, along with a rigorous training in laboratory practice, he was exposed to the camaraderie and intellectual spark of the scientists who gathered there from around the world. "This was," he says, "an attractive aspect of being a scientist."

And so, in 1994, Chang enrolled in the MD-PhD program of Harvard and MIT. After two years of medical school, he joined the laboratory of the Nobel laureate David Baltimore. In his doctoral studies, Chang explored the signaling cascades and biochemical mechanisms by which cells are chopped up from the inside during a type of programed cell death known as apoptosis, completing his PhD in just two years. After finishing medical school at Harvard in 2000, Chang returned to California for an internship at the Santa Clara Valley Medical Center followed by a residency in clinical dermatology at Stanford.

Toward the genome

Eager to enter the burgeoning field of

Chang with scientist Tansu Bagdatli.

genomics, Chang joined Brown's Stanford lab as a postdoctoral fellow in 2001, where he would make his pioneering contributions to our understanding of how cells know where they are in the body. Using DNA microarrays, which fish out the transcripts of expressed genes, Chang detailed how distinct gene expression patterns in fibroblasts reflect their locations in relation to the various axes of the body. Aside from defining the outlines of the organismal GPS, the work opened a new window into the deadliest outcome of malignancy—metastasis.

"If we're saying that different parts of the skin have beautifully laid out address codes," he explains, "a cancer cell going from one part of the body to another clearly has to deal with those address codes. I was able to characterize the gene expression profiles associated with cancer cells that have different rates and proclivities for metastasis, which turned out to be pretty useful."

Chang now became increasingly curious about the means by which so many distinct gene signatures are generated in cells. "With a few exceptions," says Chang, "cells of the body have the same DNA. But they make different choices about which genes to turn on and off. So the next question was, how does that happen?"

The first map of the human genome, completed just before Chang began his postdoctoral fellowship, had surprised everyone by its paucity of protein-coding genes, which numbered in the range of 20,000. Researchers had expected it would encode five times as many. "We were doing all these experiments probing just 2% of
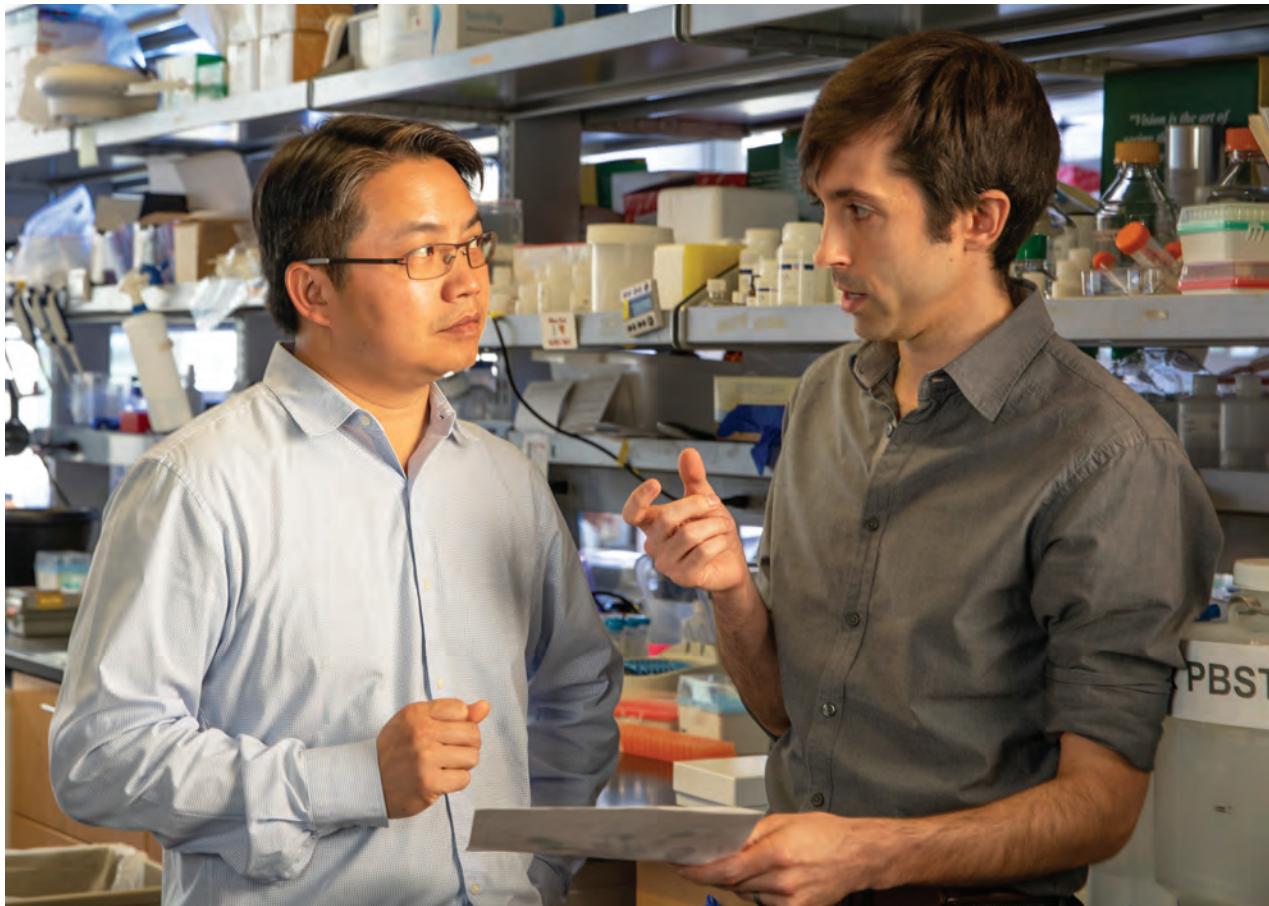
Chang and postdoc Ryan Corces.

the genome's output," says Chang. "A major theme of my work became understanding the hidden information in the noncoding genome. Subsequent work has shown that most of the variation associated with human disease resides there."

Into the dark

Chang's lab began by adapting a version of the microarray called a tiling array to look not just for mRNA transcripts of genes but for all RNAs read out of the genome. Contrary to their (and the field's) expectations, he and his colleagues saw scads of RNA transcripts emerging from regions known to be devoid of protein coding genes. These molecules, they found, belonged to a sprawling family of RNAs—now known to be some 60,000-strong—that have many of the properties of mRNAs yet do not encode proteins.

The molecules, subsequently named long noncoding RNAs (lncRNAs), turned out to be variegated in form, selectively expressed in tissues and deployed across the entire protein-coding genome. In 2007, Chang and his colleagues described in *Cell* how one lncRNA, which they later named HOTAIR, suppresses the expression of HOX genes, which dictate the body plan during development and—they discovered— the assignment of positional identity in fibroblasts.

Since discovering lncRNAs, Chang's lab has developed groundbreaking methods to harness them for the study of the genome's architecture and expression. His group has meticulously mapped lncRNA association with the genome and delineated the principles guiding those interactions. Other

studies have explored the functional role of lncRNAs, revealing how they participate in everything from embryonic development to stem cell biology to cancer. Chang and his colleagues found, for example, that HOTAIR and another lncRNA, HOTTIP, serve as scaffolds for protein complexes that chemically modify DNA and its protein packaging—collectively referred to as "chromatin"—to control HOX expression. Such "epigenetic" modification determines which genes in a given cell are switched on or off.

Stretched out, the DNA in a cell would be about two meters long. Yet it is, remarkably, crammed into a nucleus just 10 microns across. To fit, DNA is tightly spooled and packed into fractal chromatin structures that sequester most of its information from the cell's gene-reading machinery. Only DNA that must be read for a cell to survive and perform its unique function is unraveled and made available to the protein machines that control and execute gene expression.

Which segments of the genome are so favored is determined in large measure by epigenetic modifications, and these modifications are almost universally disordered in cancers. Chang's work has shown that lncRNAs are intimately involved in these processes. He and his colleagues discovered, for example, that HOTAIR reprograms chromatin to drive cancer and its metastasis. They also have mapped the lncRNAs expressed in various cancers along with the gene expression profiles associated with each.

Mapping access

By 2012, Chang's ambitions had grown to encompass the mapping and characterization of all accessible regions of the dark genome. These stretches would also include enhancers and suppressors, which are DNA sequences that produce no RNA of any kind but guide proteins to mute or amplify the expression of distant genes. To that end, Chang began collaborating with Stanford biophysicist William Greenleaf and a gifted

"A major theme of my work became understanding the hidden information in the noncoding genome.**"**

graduate student, Jason Buenrostro, who now has his own lab at Harvard University, to develop the required methods.

The two-step method they reported in *Nature Methods* in 2013, dubbed ATAC-seq, profiled the accessible genome with a million times greater sensitivity than comparable techniques, which would take days to furnish results. They showed that ATAC-Seq could, by contrast, profile the accessible chromatin of T cells overnight and from a standard clinical blood-draw. "Turning it into a daily blood test was pretty cool, we thought," says Chang.

By 2015, the researchers reported in *Nature* the development of an ATAC-Seq to profile individual cells. The study revealed that even immune cells that appear to belong to the same subclass display enormous diversity in their genomic expression, an insight of material relevance to immunotherapy. To test the method's clinical utility, the researchers picked a cancer Chang treats as a dermatologist—cutaneous T cell leukemia (CTCL), which presents in the skin and is treated with a drug that inhibits an epigenetic modification.

"Only a subset of patients benefit from this

"Almost half of the DNA elements that we found in cancer were not known to be active before in the atlas of normal tissues."

drug and we have no way of knowing who's benefiting until they've gone through multiple rounds of therapy," says Chang. "We asked, can we take blood samples from patients as they go through this treatment and use our method to watch the chromatin in real time to see what's happening?"

The researchers reported in *Cancer Cell* in 2017 that only patients whose chromatin was altered during treatment benefited from the therapy. "Those whose chromatin didn't change did not benefit," says Chang. "Their cancer cell counts did not drop." With more vetting in clinical trials, the technology could give clinicians an early warning that other treatments might be in order for a given CTCL patient.

Inspired by that success, Chang and Greenleaf decided next to similarly apply ATAC-Seq to a broad range of cancers. Chang's new effort coincided with his appointment to the Ludwig Professorship, which provided him with the resources in part to pursue this ambitious goal. "The wonderful gift of the Ludwig Institute is that we are able to quickly pursue new and exciting ideas, including high risk ideas that have the potential for big rewards," says Chang.

To examine the accessible genome across cancers, Chang and his colleagues modified ATAC-Seq so that it could be used on archival samples of tumors. This would permit the analysis of human tumor samples stored

in The Cancer Genome Atlas (TCGA), a vast collection that dates back more than a decade, is annotated with clinical information and has been exhaustively analyzed in other types of genomic studies.

"We'd know who got better, who had a worse outcome, how they responded to different drugs," says Chang. "If we could only work with fresh samples, we'd have to wait another ten years for something to happen prospectively. The TCGA samples represented this ability to go to the samples that had the most information, apply cutting edge genomic technologies and learn something new."

Their study, published in *Science* at the end of 2018, surveyed the accessibility of genomes in 410 tumor samples representing 23 types of cancer to map DNA sequences that regulate gene expression in the malignancies. By integrating these results—which identified 562,709 such "cis-regulatory elements"—with other genomic, clinical and biochemical information about the same tumors, the researchers identified such things as new molecular subtypes of cancers and their relationship to patient prognoses. Notably, the findings also shed light on how inherited variations in DNA sequence in noncoding DNA can predispose people to cancer—illuminating a poorly understood aspect of cancer risk.

Analysis of the data revealed how mutations in noncoding sequences thousands of bases away from a gene can alter chromatin to create a newly accessible stretch of DNA that promotes the aberrant expression of that gene. In a bladder tumor, for example, a mutation generates a new binding site for a protein that regulates gene expression, driving the expression of a neighboring gene that influences cell size, motility and shape—all key factors in cancer metastasis. The findings indicate that unique suites of such mutations may drive different types of cancer.
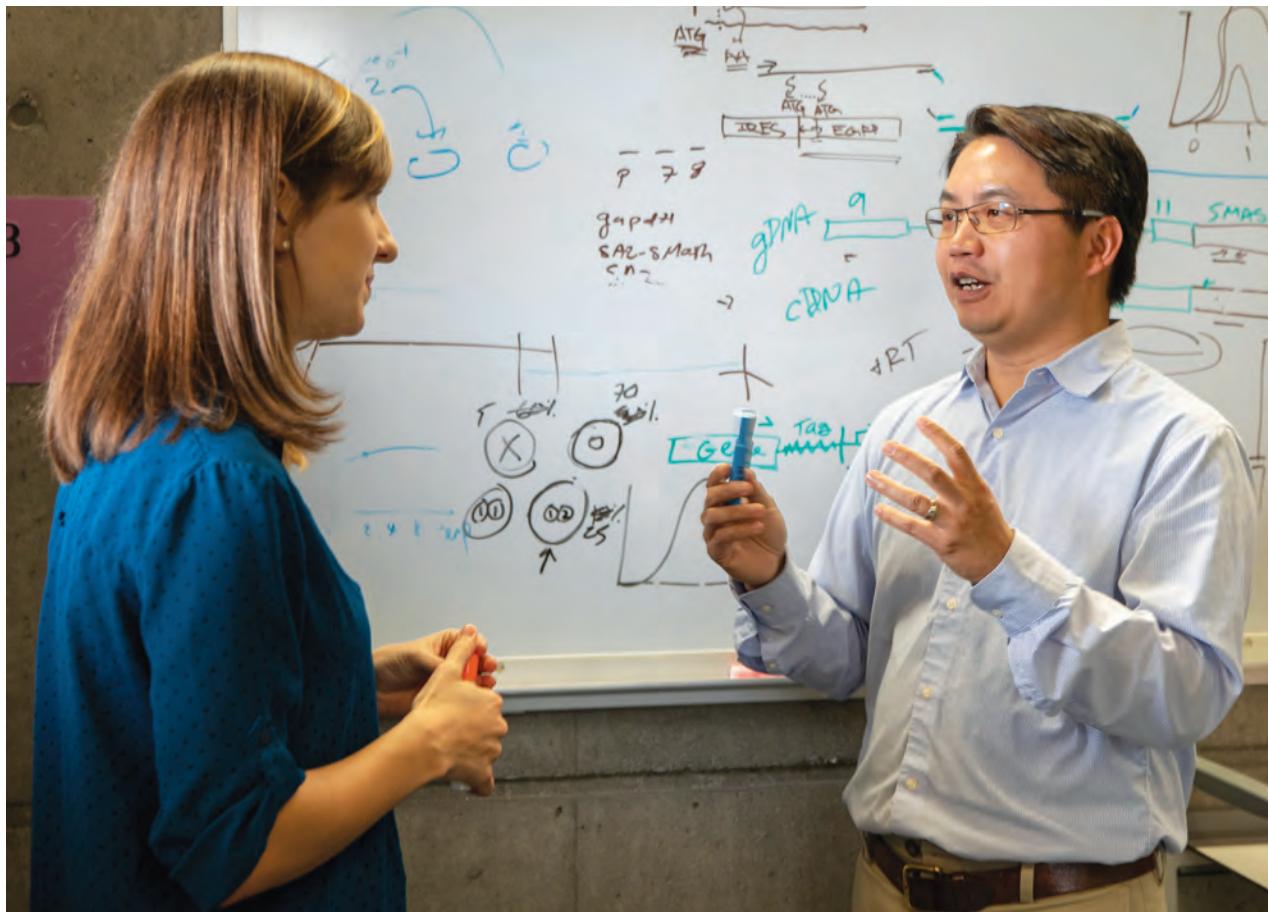
Chang and graduate student Kathryn Yost.

By layering their chromatin accessibility map over the gene expression data for various cancers, the researchers also identified tens of thousands of likely interactions between regulatory elements of DNA and genes known to play an important role in cancer and the ability of tumors to evade immune attack. This is invaluable information: Mutations to genes have consequences on proteins that can be detected and functionally analyzed. But mutations and variations in noncoding DNA sequences do not produce such readily measurable readouts, and most sequence variations associated with disease reside in just such stretches of the genome.

"Using the chromatin accessibility map, you could actually get a sense of which mutations had a biochemical consequence on the DNA element, making it more accessible or less so," says Chang. "I hope that will prove to be a useful lens for distinguishing passenger mutations that have no biochemical consequence from mutations that actually change chromatin accessibility in human cancers."

The findings, he notes, also demonstrate that the genome is every bit as complex as you'd expect it to be. "Almost half of the DNA elements that we found in cancer were not known to be active before in the atlas of normal tissues," observes Chang. "They're only accessed in the pathology of cancer, which suggests there's a lot left to be learned about the genome."

Fortunately, Chang is looking into the matter. ◢