AAAS WACHTEL PRIZE ESSAY

# A genome-wide net to catch and understand cancer

**Neville E. Sanjana**[1,2]



**Genome-scale forward genetic screens elucidate the genetic basis of therapeutic resistance, tumor evolution, and metastasis in diverse human cancers.**

Cancer is a disease of the genome. Tumors from mutagenic cancers such as melanoma can harbor up to 1 million mutations, making it challenging to understand which mutations are central to the disease and which are simply passengers (*1*). Grasping the functional consequences of different mutations is vital: Can we predict whether a patient will respond to chemotherapy or immunotherapy, or which tumors are most likely to metastasize into a more aggressive and potentially terminal disease? In my work over the past few years, I have developed high-throughput gene-editing approaches to elucidate key genes and noncoding elements involved in therapeutic resistance, in immunotherapy failure, and in metastasis of primary tumors to distal organs (*2*–*5*). The easy programmability of clustered regularly interspaced short palindromic repeats (CRISPR) nucleases has made it possible to target thousands of different locations in the human genome in virtually any tumor or cell type. This transformative technology has empowered individual investigators to answer genome-scale questions without the difficulty of acquiring a large cohort of tumor samples (Fig. 1).

## FIRST STEPS: KNOCKOUT MUTATIONS THAT DRIVE DRUG RESISTANCE IN MELANOMA

*BRAF* is the most commonly mutated gene in melanoma: 50 to 70% of melanomas carry a gain-of-function mutation in *BRAF* (*1*). In 2011, the U.S. Food and Drug Administration approved the drug vemurafenib, a targeted inhibitor of mutant BRAF that can kill these melanoma cells. In the clinic, vemurafenib extends survival by several months, but, in all cases, resistance develops over time. To pinpoint loss-of-function (knockout) mutations that cause vemurafenib resistance genome-wide, my colleague Ophir Shalem and I developed an initial CRISPR library to knock out nearly all of the ~20,000 genes in the human genome (*2*). Each gene was targeted by several distinct guide RNAs, which would guide the CRISPR nuclease Cas9 to knock out these genes. After introducing this Genome-Scale CRISPR Knock-Out (GeCKO) library into a BRAF-mutant melanoma, we treated the tumor cells with either vemurafenib or a vehicle.

We looked for genes targeted with multiple CRISPR guide RNAs that were able to grow in vemurafenib. By only considering genes with consistent results across multiple guide RNAs, we were less likely to be misled by any individual guide RNA, which may have off-target activity elsewhere in the genome. With this approach, we were able to find all previously known genes where a loss-of-function mutation results in vemurafenib resistance and also discovered several new genes that we subsequently validated (*2*).

To put these capabilities in perspective, human cells have historically not been a genetic model organism. Unlike bacteria, yeast, fruit flies, or mice, human genomes have been challenging to edit and, particularly, to do so in a sufficiently high-throughput fashion for genetic screens. It was an eye-opening moment to realize how straightforward it had become to manipulate every single gene in the human genome and how enabling CRISPR-based pooled screens could be for studying drug resistance and beyond.

## THE OTHER 98% OF THE GENOME: NONCODING DRIVERS OF DRUG RESISTANCE

Although targeting all of the ~20,000 genes in the human genome explores a large number of potential genetic drivers, the complete genome—including regions outside of genes—is an even larger place. Less than 2% of the human genome contains protein-coding sequences. Most of the 3 billion bases in our genome consist of noncoding regions, where the relationship between somatic mutations and function is even less clear (*6*). Genome-wide association studies have shown that most disease-associated variants lie in noncoding regions (*7*). With my colleague Jason Wright, I was curious to see whether we could extend our findings about genes that modulate vemurafenib resistance to regulatory regions near these genes (*3*).

To satisfy this curiosity, we designed noncoding CRISPR libraries that targeted 200 kb surrounding the protein-coding sequence of genes where we had previously established that loss-of-function mutation results in vemurafenib resistance. These saturating mutagenesis libraries densely tile over promoters, introns, untranslated regions, and intergenic space adjacent to known vemurafenib resistance genes. Overall, we targeted nearly 1 Mb of the noncoding genome and asked whether mutations in noncoding regions could phenocopy loss of function of the nearby gene. Amazingly, we found many regions where the answer is yes.

As expected, we found that well-known regulators of gene expression, such as the promoter, were sensitive to mutagenesis. That is, mutations in key regions of the promoters of genes whose loss triggers resistance resulted in a large increase in drug resistance. In addition, a greater number of regulatory elements were found proximal to coding regions, and the 5′ side of these genes contained more regulatory elements than the 3′ side (*3*). Both results are in agreement with large surveys of genetic variation (*8*). Surprisingly, we also found several regions distal to genes, where mutations also triggered drug resistance. These regions tended to be enriched in open chromatin to facilitate transcription factor binding and displayed histone modifications associated with enhancers (such as H3K27Ac). After mutagenesis, we could measure decreases in transcription factor binding that coincided with decreased gene expression. We also found remodeling of histone posttranslational modifications to reflect loss of enhancer or promoter activity.

[1]New York Genome Center, New York, NY 10013, USA. [2]Department of Biology, New York University, New York, NY 10003, USA. Email: neville@sanjanalab.org

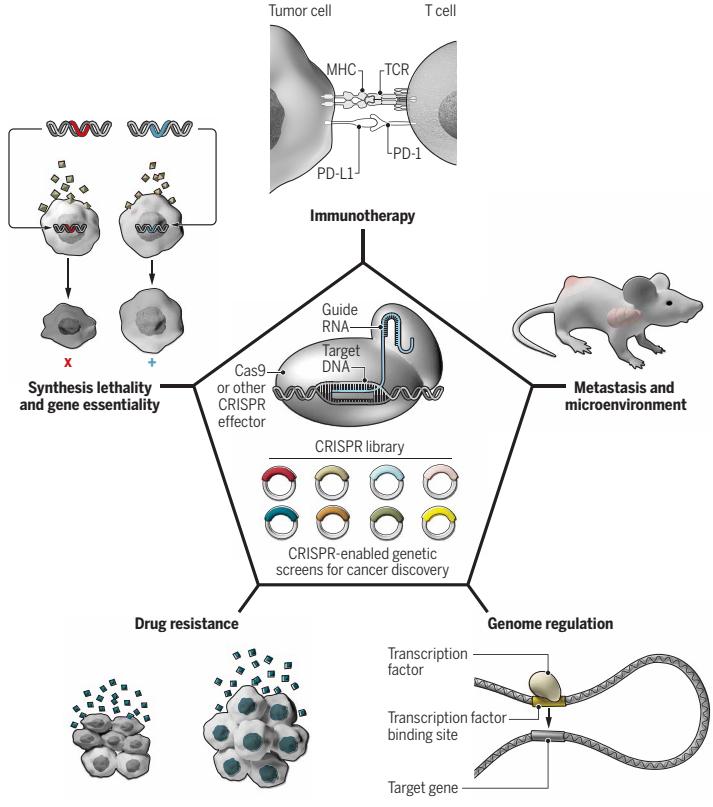
Using chromosome conformation capture to map three-dimensional interactions, we found that regions enriched in functional elements identified in the CRISPR screen tended to contact the resistance gene promoter even when their linear (chromosomal) distance from the promoter was large.

## METASTASIS: ADAPTING CRISPR SCREENS FOR IN VIVO APPLICATIONS

The ability to rapidly interrogate all genes in the genome or large regions of the noncoding genome to find drivers of drug resistance is valuable, but many properties of cancer evolution cannot be replicated in cell culture. One of the most important hallmarks of cancer progression is metastasis of a primary tumor to distal organs. As a first step to understanding metastasis, O. Shalem and I designed updated GeCKOv2 genome-scale libraries for human and for mouse (*9*). Due to the increasing popularity of CRISPR-based functional genomics, these reagents were widely distributed. For example, in 2015, our lentiCRISPRv2 vector was the most requested plasmid worldwide from Addgene (*10*).

Using this new mouse genome–wide library, my colleague Sidi Chen and I developed a genome-scale in vivo CRISPR screen to understand the key drivers of metastasis (*4*). We transduced a KRAS- and TP53-mutant lung cancer cell line with the GeCKOv2 mouse library ex vivo and then transplanted these cells subcutaneously into nude mice. After several weeks, the library-transduced cells had metastasized to the lungs, whereas mice receiving cells with Cas9 alone did not develop metastases. To find genes where loss of function increased the cells' metastatic ability, we compared the differential representation of guide RNAs in metastases and primary tumors.

In this experiment, we identified known tumor suppressors such as PTEN and NF2 and several new genes and microRNAs that selectively increased the metastatic capability of the cancer. Because we simultaneously measured proliferation both at the primary tumor site and at distal metastatic sites, we could compute a relative abundance ratio for each gene. By doing so, we teased apart genes that were selectively enriched in metastases and those that promoted proliferation of the primary tumors but did not metastasize. Overall, we found that the effect of loss-of-function mutations on metastasis strongly correlates with abundance in late-stage primary tumors.




CREDIT: A. KITTERMAN/*SCIENCE TRANSLATIONAL MEDICINE*

**Fig. 1. CRISPR-enabled functional screens have accelerated genomic discovery in cancer.** Pooled libraries of CRISPR guide RNAs (colored circles) can target all ~20,000 protein-coding genes in human and mouse genomes. CRISPR libraries can pinpoint genetic drivers of cancer metastasis, genes involved in therapeutic resistance to chemotherapies and immunotherapies, and genes essential for tumor growth. In addition, CRISPR libraries targeting noncoding regions of the genome can identify key regulatory elements that modulate those genes. MHC, major histocompatibility complex; TCR, T cell receptor; PD-L1, programmed death-ligand 1; PD-1, programmed death 1.

## NEW FRONTIERS IN T CELL IMMUNOTHERAPY AND TWO CELL–TYPE SCREENS

In vivo genetic screens can capture the complexity of the tumor microenvironment in ways that are not possible in vitro. However, this complexity can be a double-edged sword because it is difficult to isolate interactions between specific sets of cells in vivo. One particularly relevant tumor-interacting cell is the T cell, a target in cancer immunotherapy. Immunotherapy, which harnesses the body's immune system to fight cancer, has blossomed over the last decade because of the durable response rates seen in many different cancers. However, not all patients respond to immunotherapy.

With my colleague Shashank Patel, we wondered whether we could map out genetic drivers of resistance to T cell immunotherapy (*5*). We developed a two cell–type CRISPR screen to examine interactions between cytotoxic T cells and melanoma cells. After introducing the GeCKOv2 library into human melanoma cells, we challenged these cells with a near-lethal dose of cytotoxic T cells engineered with a T cell receptor that detects the NY-ESO-1 antigen displayed by the tumor cells. The top two gene hits in the screen were HLA and B2M—the two components of the major histocompatibility complex that displays antigens on the cell surface. Patient tumors with B2M mutations are refractory to immunotherapy. Other key players in antigen processing and presentation were also highly enriched, along with several genes not previously implicated in therapeutic resistance. To understand the clinical relevance of these findings, we performed a meta-analysis of exome sequencing data from patient tumors before treatment with immune checkpoint antibodies and found that nonresponders tended to harbor mutations in our top-ranked genes. This suggests that the data from this two cell–type screen may help to predict which tumors would not respond to immunotherapy.

We have only begun to understand the functional importance of different mutations in cancer, and I am very grateful to the

collaborators and mentors who have included me in this wonderful journey. Going forward, new gene-editing technologies and computational methods will be needed to better understand how mutations affect cancer evolution and therapeutic resistance. My hope is that, in the not-too-distant future, clinicians can interpret patient-specific tumor mutations in coding and noncoding regions to precisely tailor treatments from which cancers cannot escape.

## REFERENCES AND NOTES


1. N. K. Hayward, J. S. Wilmott, N. Waddell, P. A. Johansson, M. A. Field, K. Nones, A.-M. Patch, H. Kakavand, L. B. Alexandrov, H. Burke, V. Jakrot, S. Kazakoff, O. Holmes, C. Leonard, R. Sabarinathan, L. Mularoni, S. Wood, Q. Xu, N. Waddell, V. Tembe, G. M. Pupo, R. De Paoli-Iseppi, R. E. Vilain, P. Shang, L. M. S. Lau, R. A. Dagg, S.-J. Schramm, A. Pritchard, K. Dutton-Regester, F. Newell, A. Fitzgerald, C. A. Shang, S. M. Grimmond, H. A. Pickett, J. Y. Yang, J. R. Stretch, A. Behren, R. F. Kefford, P. Hersey, G. V. Long, J. Cebon, M. Shackleton, A. J. Spillane, R. P. M. Saw, N. López-Bigas, J. V. Pearson, J. F. Thompson, R. A. Scolyer, G. J. Mann, Whole-genome landscapes of major melanoma subtypes. *Nature* **545**, 175–180 (2017).
2. O. Shalem, N. E. Sanjana, E. Hartenian, X. Shi, D. A. Scott, T. S. Mikkelsen, D. Heckl, B. L. Ebert, D. E. Root, J. G. Doench, F. Zhang, Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84–87 (2014).
3. N. E. Sanjana, J. Wright, K. Zheng, O. Shalem, P. Fontanillas, J. Joung, C. Cheng, A. Regev, F. Zhang, High-resolution interrogation of functional elements in the noncoding genome. *Science* **353**, 1545–1549 (2016).
4. S. Chen, N. E. Sanjana, K. Zheng, O. Shalem, K. Lee, X. Shi, D. A. Scott, J. Song, J. Q. Pan, R. Weissleder, H. Lee, F. Zhang, P. A. Sharp, Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. *Cell* **160**, 1246–1260 (2015).
5. S. J. Patel, N. E. Sanjana, R. J. Kishton, A. Eidizadeh, S. K. Vodnala, M. Cam, J. J. Gartner, L. Jia, S. M. Steinberg, T. N. Yamamoto, A. S. Merchant, G. U. Mehta, A. Chichura, O. Shalem, E. Tran, R. Eil, M. Sukumar, E. P. Guijarro, C.-P. Day, P. Robbins, S. Feldman, G. Merlino, F. Zhang, N. P. Restifo, Identification of essential genes for cancer immunotherapy. *Nature* **548**, 537–542 (2017).
6. A. Montalbano, M. C. Canver, N. E. Sanjana, High-throughput approaches to pinpoint function within the noncoding genome. *Mol. Cell* **68**, 44–59 (2017).
7. M. T. Maurano, R. Humbert, E. Rynes, R. E. Thurman, E. Haugen, H. Wang, A. P. Reynolds, R. Sandstrom, H. Qu, J. Brody, A. Shafer, F. Neri, K. Lee, T. Kutyavin, S. Stehling-Sun, A. K. Johnson, T. K. Canfield, E. Giste, M. Diegel, D. Bates, R. S. Hansen, S. Neph, P. J. Sabo, S. Heimfeld, A. Raubitschek, S. Ziegler, C. Cotsapas, N. Sotoodehnia, I. Glass, S. R. Sunyaev, R. Kaul, J. A. Stamatoyannopoulos, Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195 (2012).
8. GTEx Consortium, Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
9. N. E. Sanjana, O. Shalem, F. Zhang, Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784 (2014).
10. T. Ford, Your Top Requested Plasmid in 2015, *Addgene Blog* (2015); http://blog.addgene.org/your-top-requested-plasmid-in-2015.


**Acknowledgments:** I am grateful to the collaborators and mentors who have included me in this wonderful journey, including D. Altschuler, P. Sharp, and N. Restifo. In particular, I am indebted to my postdoctoral mentor, F. Zhang, who not only created a wonderful, stimulating laboratory environment but also gave me the freedom to explore different ideas. **Funding:** I received funding from the NIH through the National Human Genome Research Institute (K99/R00-HG008171 and DP2-HG010099), a Simons Center for the Social Brain Fellowship, and a Sidney Kimmel Scholar Award. **Competing interests:** N.E.S. is a co-inventor on patents relating to the described work.

10.1126/scitranslmed.aat8288

**Citation:** N. E. Sanjana, A genome-wide net to catch and understand cancer. *Sci. Transl. Med.* **10**, eaat8288 (2018).

Science Translational Medicine

# A genome-wide net to catch and understand cancer

Neville E. Sanjana

*Sci Transl Med* **10**, eaat8288.
DOI: 10.1126/scitranslmed.aat8288