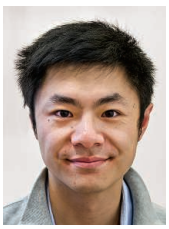
PRIZE ESSAY



FINALIST: GENOMICS AND PROTEOMICS

Tim Wang
Tim Wang received his undergraduate degree from the University of California, Berkeley, and his Ph.D. from the Massachusetts Institute of Technology. He is currently investigating the structural and functional organization of neocortical circuits as a postdoctoral fellow at the Janelia Research Campus in the laboratory of K. Svoboda.
www.sciencemag.org/content/362/6417/904.3

GENOMICS AND PROTEOMICS

Paring down to the essentials

A genome-wide library of single gene knockouts enables high-throughput functional analysis of mammalian genes

By **Tim Wang**

In the early 1950s, television sets were complicated and expensive, putting them beyond the reach of most American households. To lower prices, businessman and engineer Earl "Madman" Muntz devised an ingenious solution (*1*). Armed with a pair of snippers, he would take a competitor's product and proceed to pluck off component after component until, inevitably, the device would break. At this point, he would put back the last piece removed and thereby restore function. In the end, he was left with a minimal design that contained only the essential circuit elements, which could then be manufactured at a reduced cost.

This practice of "Muntzing" is not only a clever way to dissect electronic circuitry but also a powerful method by which to probe the genetic circuitry of biological organisms. By disrupting individual genes (either at random or in a systematic fashion) and observing the resultant phenotypes, one can efficiently identify the set of genes underlying any cellular pathway.

In several model organisms, large-scale genetic screens have provided insights into fundamental biological processes, including the cell cycle (*2*), programmed cell death (*3*), and embryonic development (*4*). In mammalian cells, however, the development of similar techniques has lagged behind, leaving our understanding of human genes incomplete.

As a graduate student working jointly in David M. Sabatini's laboratory at the Whitehead Institute and Eric Lander's laboratory at the Broad Institute, I set out to develop a general strategy for conducting genome-wide loss-of-function screens in human cells (see the figure). To accomplish this task, I used the clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 system, which functions as an adaptive immune system in bacteria [reviewed in (*5*, *6*)]. During my second year in the laboratory, two groups had reported a method for mammalian genome editing via heterologous expression of two components of the CRISPR pathway: a Cas9 nuclease that is directed to cut specific genomic sequences by means of a single-guide RNA (sgRNA) (*7*, *8*).

To develop a platform that would be amenable for large-scale genetic screening, I designed a library of sgRNAs that targets all protein-coding genes in the human genome (*9*). This library could be introduced into any cultured cell line to generate a collection of "knockout" mutants in which a single gene is inactivated in each cell. Using high-throughput sequencing, the abundance of each mutant in the population can be tracked during the course of an experiment.

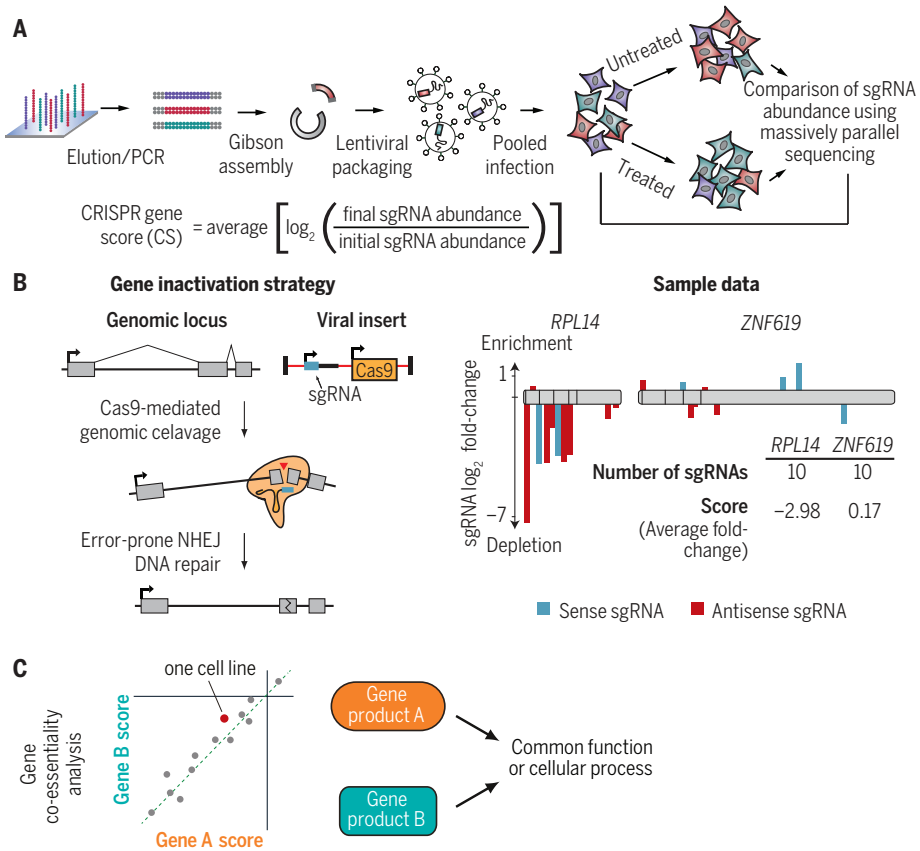
In an initial study, I applied this technique to search for genes that are necessary for human cells to grow and divide in culture (*10*). This survey produced a provisional catalog of all cell-essential genes. As a whole, these genes show broad conservation across species, are expressed at high levels, and contain few inactivating polymorphisms in the human population. The majority of these genes encode well-defined components of fundamental cellular pathways (such as ribosomal subunits).

Surprisingly, this study also uncovered a large group of genes (more than 300) for which no molecular function had been previously ascribed. For this set of genes, bioinformatics analysis and comparison to proteomic datasets revealed substantial enrichment in proteins found in the nucleolus and those containing domains related to RNA processing. Indeed, directed studies confirmed key roles for three of these genes in various stages of posttranscriptional RNA processing. More broadly, these findings indicate that the molecular components of many critical cellular processes in mammalian cells have yet to be fully defined.

Next, I profiled gene essentiality across a much larger collection of human cancer cell lines derived from acute myeloid leukemias, which are the most common form of adult leukemia (*11*). Whereas most genes were either essential or dispensable in all lines, a small fraction showed variable essentiality across the cell lines. This latter class of genes provided two major avenues for further investigation.

Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA. Email: wangt@janelia.hhmi.org

PHOTO: CEAL CAPISTRANO



CRISPR/Cas9–based genetic screening. (**A**) Outline of sgRNA library construction and genetic screening strategy. Briefly, sgRNAs are synthesized on a microarray, cloned, and packaged into lentiviral particles that can be used to transduce a cell line of interest to generate a mutant library. Cells can be cultured under various conditions, and the fold-change in abundance between the initial seeding and final collection of each mutant can be quantified by means of high-throughput sequencing. (**B**) (Left) Gene inactivation through Cas9-mediated genomic cleavage is directed by a 20–base pair sequence at the 5′ end of the sgRNA. Lesions repaired by the error-prone nonhomologous end joining (NHEJ) pathway frequently result in nonfunctional, frame-shifting mutations. (Right) sgRNAs that target essential genes are depleted from the population, whereas those that target dispensable genes are maintained, as illustrated from the data of two neighboring genes: *RPL14*, an essential ribosomal protein gene, and *ZNF619*, a dispensable gene that encodes a zinc finger protein. (**C**) Strategy for identifying functionally related sets of genes by using the gene essentiality dataset in (*11*). Because genes acting in the same cellular pathway should show similar patterns of essentiality across cell lines, functional gene networks can be mapped through correlation-based analysis. [Adapted from (*9*–*11*)]

First, I used the "gene essentiality" dataset to pinpoint genetic dependencies in particular cancer subtypes. In contrast to core essential genes, variably essential genes are attractive targets for anticancer therapies because their inhibition is less likely to be broadly cytotoxic.

Because the genomes of these cell lines had already been characterized, it was possible to relate patterns of gene essentiality with the presence of specific mutations. This approach could not only identify the mutant oncogene driving the proliferation of each line but, importantly, could also detect vulnerabilities in genes that were not mutated (and thus not detectable with genome sequencing). For example, I uncovered several liabilities associated with oncogenic *Ras*, a frequently mutated oncogene that has proven difficult to inhibit directly (*12*). In the future, gene essentiality profiling of cancer cell lines, and even patient tumor cells, could guide personalized cancer treatment and allow for the discovery of previously unidentified gene targets for therapeutic intervention.

Second, the dataset was used to map functional gene networks. I noticed that genes acting in the same cellular pathway tended to show similar patterns of essentiality across cell lines. This observation suggested that functionally related sets of genes can be recognized by analyzing correlations in gene essentiality. By applying a "guilt-by-association" approach, genes were systematically clustered by function, which in turn revealed new gene relationships, the essential substrates of enzymes, and the molecular functions of uncharacterized proteins.

In a number of collaborative studies, CRISPR/Cas9–based screens have been performed to uncover the molecular underpinnings of diverse biological processes. Screens in a $CD4^+$ cell line, for example, have identified host factors that are required for HIV infection (*13*). Genetic analysis of cells cultured under limiting nutrient conditions or in the presence of chemical inhibitors have yielded key insights into mitochondrial function (*14*, *15*).

Others have greatly extended this CRISPR-based approach to examine noncoding regions of the genome (*16*), to dissect pairwise gene interactions (*17*), to perform gain-of-function experiments (*18*), and to analyze richer phenotypes by use of single-cell RNA sequencing (*19*).

Taken together, these efforts have enabled genetic analyses of mammalian cells to be conducted with a degree of rigor and completeness previously possible only in microorganisms. With further advances in the design of sgRNA libraries, Cas9 effector proteins, and readout strategies, it is tremendously exciting to imagine what discoveries will be made by "Muntzing" around in the cell. ■

REFERENCES AND NOTES

1. R. C. Post, *Technol. Culture* **46**, 773 (2005).
2. L. H. Hartwell, J. Culotti, B. Reid, *Proc. Natl. Acad. Sci. U.S.A.* **66**, 352 (1970).
3. H. M. Ellis, H. R. Horvitz, *Cell* **44**, 817 (1986).
4. C. Nüsslein-Volhard, E. Wieschaus, *Nature* **287**, 795 (1980).
5. L. A. Marraffini, E. J. Sontheimer, *Nat. Rev. Gen.* **11**, 181 (2010).
6. P. Horvath, R. Barrangou, *Science* **327**, 167 (2010).
7. L. Cong *et al.*, *Science* **339**, 819 (2013).
8. P. Mali *et al.*, *Science* **339**, 823 (2013).
9. T. Wang *et al.*, *Science* **343**, 80 (2014).
10. T. Wang *et al.*, *Science* **350**, 1096 (2015).
11. T. Wang *et al.*, *Cell* **168**, 890 (2017).
12. A. D. Cox *et al.*, *Nat. Rev. Drug Discov.* **13**, 828 (2014).
13. R. J. Park *et al.*, *Nat. Gen.* **49**, 193 (2017).
14. K. Birsoy *et al.*, *Cell* **162**, 540 (2015).
15. N. Kory *et al.*, *Science* **362**, 791 (2018).
16. N. E. Sanjana *et al.*, *Science* **353**, 1545 (2016).
17. F. J. Najm *et al.*, *Nat. Biotechnol.* **36**, 179 (2018).
18. S. Konermann *et al.*, *Nature* **517**, 583 (2014).
19. B. Adamson *et al.*, *Cell* **167**, 1867 (2016).

ACKNOWLEDGMENTS

The author thanks R. A. Saxton, E. S. Lander, and D. M. Sabatini for reviewing this essay and the Howard Hughes Medical Institute for funding. **Competing interests:** The author is an inventor on a patent for functional genomics using CRISPR/Cas (U.S. patent application 15/141,348) and a scientific cofounder of KSQ Therapeutics, which is using CRISPR-based genetic screens to identify to drug targets.

10.1126/science.aav6872

PHOTO: N. DESAI/*SCIENCE*, BASED ON (*9*–*11*) AND TIM WANG

Paring down to the essentials

Tim Wang

Science **362** (6417), 904.
DOI: 10.1126/science.aav6872