20 years of *Nature Biotechnology* research tools

Anna Azvolinsky, Laura DeFrancesco, Emily Waltz & Sarah Webb

Authors of some of the most highly cited *Nature Biotechnology* papers that describe research tools discuss their work and challenges for their fields.

One way of defining biotech is as a source of new research tools for probing biology and providing new insights into the workings of living cells and organisms. The following article provides a sampling of a set of research tools across the full span of our editorial scope. Readers are also referred to other Features in this issue on biomedical (p 262) and nonbiomedical applications (p 267).

As *Nature Biotechnology* is a journal of bioengineering, it is not surprising that the collection of papers highlighted here is dominated by tools for molecular engineering and genomics. Gene editing tools, in particular CRISPR-Cas9, spawned of late a spate of papers in our pages and elsewhere; you'll find a suite of papers demonstrating the use of these RNA-guided enzymes in several model systems. As high-throughput technologies, such as next-generation sequencing, have ramped up in speed, scaled down in cost and become ever more sensitive—even down at the level of single cells—large data sets are accumulating in public repositories. As a result, we have chosen to highlight several papers providing tools for sorting, assembling and interrogating massive data sets.

This being our third such effort—we profiled similar collections in our tenth and fifteenth anniversary issues—we chose to highlight only papers not previously profiled. A short summary of other papers on research tools from our tenth and fifteenth anniversary issue Features can be found in **Box 1**; the complete collection of articles can be found here[1,2]. We acknowledge that the papers below provide a limited snapshot of all the research that has appeared in our pages, and we thank the entire research community for its contributions to the journal over all these years.

Anna Azvolinsky, Emily Waltz and Sarah Webb are freelance writers. Laura DeFrancesco is Senior Editor at Nature Biotechnology.

How to assemble a transcriptome

Since it was first described in 2008 (refs. 3–6), high-throughput sequencing of RNA (RNA-seq) has eclipsed DNA microarray analysis as the method of choice for quantifying gene expression. It offers not only a larger dynamic range than microarrays, but also an ability to discover new transcripts without any prior knowledge of sequence. One key reason for the wide adoption of this approach was the dissemination of software tools for assembling short reads from the sequencer into full transcripts. And two papers in *Nature Biotechnology* made important contributions: Cufflinks became the method of choice for those analyzing transcripts in model organisms with a complete reference genome[7]; the other, Trinity, opened up RNA-seq to organisms lacking a reference genome[8].

In 2007, Lior Pachter of the University of California, Berkeley, literally heard opportunity knocking on his door in the form of a visiting computer science graduate student who stopped by his office and offered his time and effort for a year. The student, Cole Trapnell, then at the University of Maryland College Park, had experience aligning large numbers of genomic sequencing reads to their place in the genome. Pachter had been studying conservation of genome structure by comparing the alignment of homologous genes among various species. The two realized they were both thinking about the future of sequencing and, particularly, the then as-yet unpublished technique of RNA-seq. Foreseeing the needs of biologists, the two began working on ways to align the 25-base-pair-long reads generated at that time by the technology. "We quickly realized that what we needed to do was to assemble transcripts from reads to characterize transcriptomes and to figure out the abundance of transcripts, the question that was circulating in the minds of the researchers developing RNA-seq assays," says Pachter.

The end result was a 2010 *Nature Biotechnology* paper describing Cufflinks[7] a computer program that assembles and quantifies transcripts using RNA-seq data and allows comparisons between samples. "It's a foundational paper that has been hugely impactful," says Michael Snyder, at Stanford University (Stanford, CA, USA), who was not involved in the study. "It is certainly one of the first and most robust tools for quantitative analyses of gene expression." One advantage of the program is its ability to discover new transcripts, including alternatively spliced RNA variants using a reference genome to align the sequences. "It was a powerful tool because it allowed biologists to make discoveries of new genes and new transcripts directly from the data," says Pachter.

Working along with some RNA-seq developers, the team instilled features in the program—even before the papers on RNA-seq had come out—that could evolve with the technology, including an increase in the length of RNA-seq reads. By the time the paper appeared, many researchers were already using RNA-seq technology. The tool made the data analysis accessible to lots of people without a computational background, according to Snyder.

Trapnell, Pachter and their colleagues have since published a 'Cufflinks 2.0' version[9] to analyze bulk data from multicondition

Box 1 Previously, in our hall of fame

Previous incarnations of this Feature have highlighted many other research tools published in our pages. Below, we provide brief summaries of these studies and their advances. For further details, readers are referred to the originals[1,2].

- **1996.** David Lockhart's group at Affymetrix (Santa Clara, CA, USA) described the first gene expression microarray containing 130,000 oligonucleotide probes and used it to profile 118 genes[31].
- **1996**. Sangi Tyagi and Fred Kramer described molecular beacons, a new kind of probe for quantifying nucleic acids in real time and without labels[32].
- **1999**. Ruedi Aebersold developed a method for quantifying proteins in complex mixtures, called ICAT for isotope-coded affinity tags[33].
- **2005.** Charles Lieber described a nanowire sensor with femtomolar sensitivity[34].
- **2007**. Christine Vogel published a new method for determining absolute protein levels, called APEX (absolute protein expression)[35].
- **2009**. George Church and Kun Zhang in separate papers described ways to do comparative epigenome profiling[36,37].

experiments and those with many replicates. Always trying to stay several steps ahead and to improve, Pachter's laboratory is now focused on faster ways to analyze RNA-seq data to avoid some of the cumbersome calculation steps of Cufflinks. "There are next generation, more statistically sophisticated methods coming out soon that will be suitable for much higher throughput experiments", says Pachter.

A problem that Cufflinks couldn't answer was how to map reads to transcripts for organisms with incomplete or nonexistent genomes. In 2010, the description of a *de novo* assembly platform called Trinity provided a solution[8]. Trinity was the culmination of work in Aviv Regev's laboratory at the Broad Institute of MIT and Harvard (both in Cambridge, MA, USA) and Nir Friedman's group at the Hebrew University (Jerusalem). What started as a triumvirate of computational biologists has now evolved into a large network of researchers from around the globe who work with and on Trinity to continually improve it.

The three drivers of the development of Trinity—Manfred Grabherr (previously at the Broad Institute, currently at Uppsala University, Sweden), Brian Haas of the Broad Institute and Moran Yassour of the Broad Institute and Hebrew University—broke the problem up into three parts, which they developed in parallel into a modular system. The first part, Inchworm, crawls along the reads until it finds some overlapping sequences and stitches them together, squishing down the masses of data into a manageable collection of contigs representing the unique regions of transcript isoforms; the second part, Chrysalis, clusters related contigs and constructs graphs displaying the transcriptional complexity of genes; and the third module, Butterfly, defines paths in the graphs best supported by the reads and yields a fully reconstructed transcriptome.

According to Regev, who, along with Friedman, is described by their collaborators as the glue that brought this group together and kept the project moving with an unwavering belief that it could be done, the three principals came to the project with specific research problems in mind. "They were each hard-core computational researchers with strong algorithmic minds so they came up with a great solution together. Beautifully," says Regev.

Trinity's broad appeal comes in part from the design. The modular nature lends itself to improvement—as new algorithms are designed, they can and have been swapped into the appropriate module, according to Regev. In addition, from the start the developers had in mind to make the platform usable by researchers in all kinds of settings, according to Grabherr. "[We said] let's reduce the number of parameters that you have to set and make it as easy as possible," he says.

Outreach and community building have also contributed to the platform's success and borne fruit. The next paper on Trinity[10] had authors from several dozen laboratories, and discussion groups and mailing lists have formed with users from around the globe. Questions posted on the board are answered within hours, according to Grabherr. He recalls one enthusiastic user informing the group he was going to assemble a billion reads. "We said, 'Good luck'. And what happened? He became one of the lead developers after the [*Nature Biotechnology*] publication," he says.

Steven Salzberg, of Johns Hopkins' Center for Computational Biology (Baltimore), who collaborates on a conifer genome project that involves analyzing some of the largest genomes known, recommends Trinity to his collaborators working on the transcriptome. *De novo* assembly is a problem for systems like this, where they can find upwards of 100,000 transcripts. "Trinity was basically trying to solve a very hard problem, and it was the best solution," he says.

Brian Haas and his collaborators around the world continue to improve the Trinity platform for assembling transcriptomes.

Regev waxes poetic about the Trinity project. "I am so proud of this group and especially Brian Haas' leadership." Haas is spearheading ongoing improvements to Trinity—getting it ready for super long-read integration and supercomputing infrastructures that are in the pipeline. "To feel that something you originally developed because you had a research question in your mind is now used by [a] very large number of individuals—it's just heartwarming," Regev says.

Predictive metagenomics

Characterizing diverse microbial species within complex environments like the human microbiome or soil is typically done by looking at the sequence of the 16S ribosomal RNA (rRNA) genes. Then researchers can draw inferences from the species they find and construct a qualitative picture of the role each microbe might be playing in their communities, explains Morgan Langille, of Dalhousie University (Halifax, Nova Scotia, Canada). To get a more accurate picture of the metabolic processes and other functions of diverse species within a community requires more thorough and costly—less generally accessible—metagenomic sequence analysis.

Enter PICRUSt (phylogenetic investigation of communities by reconstruction of unobserved states), a computational tool that uses 16S rRNA or other marker genes from an environmental sample to quantitatively predict what a metagenomic analysis would likely reveal[11]. "Anyone who does a 16S analysis can use this to retrieve quantitative

predictions," says Langille, who developed the program along with Jesse Zaneveld of Oregon State University (Corvallis) and their colleagues.

The University of California (UC) Davis's Jonathan Eisen says he had a Eureka moment when he first saw the paper. "It's a genius piece of work, actually. [The authors] had a nice computational method for doing what was staring the community in the face. People had talked about it, and there were a few anecdotal examples, but no one had done it before."

Starting with the identities of the microbes revealed by the 16S rRNA sequence, PICRUSt uses existing genome databases to infer each community member's function. In the simplest cases, the species might already be represented in a database. But when it's not, the program infers function in two ways: by identifying closely related species using a phylogenetic tree of the 16S rRNA gene data, or by reconstructing a phylogenetic ancestral state, which models the gain or loss of genes based on the 16S tree and previously sequenced genomic information. The predictions are done gene by gene, resulting in final predicted genomes which are then combined to create the total gene content of the bacterial community.

The tool is quantitative says Langille, because the copy number of each gene, and thus, its abundance is calculated based on orthologous genes of the nearest sequenced species neighbors. "PICRUSt makes predictions on the species level and then builds that up into a community profile," he says.

As Langille was developing the program, he and his colleagues had doubts that the tool would work because 16S sequences don't perfectly match with function and because bacteria regularly swap genes by lateral gene transfer. "I think many people didn't think it would work, but, to our own surprise, it worked really well," he says.

Morgan Langille himself didn't think that 16S RNA genes alone would be as informative as they turned out to be.

But, like any prediction tool, it has its limitations. The results are only as good as the reference bacterial genomes available in public databases. In addition 16S rRNA gene sequencing using PCR does not provide a perfect picture of the entire community, according to Eisen. It captures only abundant organisms while missing rare species and doesn't capture at all organisms whose 16S genes are not readily identifiable or missing entirely as in the case of viruses.

Lagille and Zaneveld continue to improve PICRUSt. On the technical side, adding more genomic sequences is improving the program. The current version treats each 16S sequence as a separate data point but the two researchers are also trying to make better functional predictions by incorporating additional environmental context information, such as the presence or absence of other species within the sample. Says Lagille, "Organisms within the same species may have very different functional capabilities depending on their environment and nearest microbial neighbors. Co-occurrence or absence of a species could give us clues about what specific organisms are doing."

Optogenetics in peripheral nerves

Studying and treating pain has been slowed by technical limitations, but an approach published in these pages in 2014 offers a potential alternative that may help. In that paper, researchers from Stanford University described how optogenetics could control pain circuitry. They used a viral vector to introduce light-sensitive proteins into pain-signaling neurons in mice, and manipulated the neurons by shining light onto the paws of awake, freely moving mice. In doing so, the researchers were able to tune the animals' pain thresholds up and down[12].

The approach was an early demonstration of optogenetics in the peripheral nervous system, and may hold promise in developing treatment for pain. "It is very exciting to see applications of optogenetics in the peripheral nervous system," says Edward Boyden, a neuroscientist at MIT (Cambridge, MA, USA) who did not collaborate on the paper. Such work can help "map out the kinds of neural circuits that could contribute to pathology outside the central nervous system," he says.

The Stanford group showed that it was possible to optogenetically manipulate peripheral pain with no implant, and in nontransgenic mice. By targeting nociceptors—peripheral sensory neurons that transmit pain—light shined on the skin of mice could reach the nerve ending just beneath the surface. And using an adeno-associated virus serotype-6 (or AAV6) to express opsin in nociceptors, they eliminated the need to generate transgenic mouse lines.

The group worked with two standard opsins: the stimulatory channnelrhodopsin-2 (ChR2), which responds to blue light, and the inhibitory halorhodopsin (eNpHR3.0), which responds to yellow light. "We were able to get expression of the opsins throughout these long pain neurons," which was a first, says Kate Montgomery, a co-author on the paper who was a graduate student in Scott Delp's neuromuscular biomechanics lab at Stanford, and is now a lead R&D scientist at Delp's spin-out, Zebra Medical Technologies (Mountain View, CA, USA). The team was surprised to find that halorhodopsin reduced pain sensitivity in mouse models of chronic neuropathic pain. "That was unexpected," says Shrivats Iyer, a graduate student in the same laboratory, who co-authored the paper.

Kate Montgomery, who was part of a team working on reducing peripheral pain with light, is commercializing the technology.

The technique also offers researchers a high-throughput way to screen opsins. "It's a rapid first line of testing. You can see within a couple of weeks after injecting a virus if the opsin is going to work *in vivo* or not," says Montgomery. "You don't have to spend a year creating a new mouse line only to discover the opsin doesn't express the way you expected, or doesn't respond to light in the way you expected."

There are limitations to the method. An external light source's reach is limited to nerve endings in the skin and has only been demonstrated in small animals. And opsin expression levels in nociceptors haven't been shown to be stable over long time periods. But for researchers interested in new ways to control pain perception, or looking for a way to rapidly test new opsins, a tool like this might come in handy.

Proteomics to the Max!

When Jürgen Cox started working on proteomics, he was an outsider to the field. With a PhD in statistical particle physics from the Massachusetts Institute of Technology (Cambridge, MA, USA), he'd only been working on biological problems for a few years, mostly on mRNA analysis for Genedata, a biotech company in Martinsried, Germany. But he started working on mass

spectrometry for metabolomics research and that led him to start thinking about the problem of identifying and quantifying proteins in proteomics research. At the time, analyzing mass spectra involved a lot of painstaking manual work, Cox says. "The computational tools that were available then—they were just not reliable enough to do this automatically."

Cox wanted to move back into academia to explore these questions. When proteomics expert Matthias Mann moved to the nearby Max Planck Institute for Biochemistry (MPI), Cox met with him and soon took a position as a postdoctoral researcher in 2006.

The collaboration led to MaxQuant—proteomics data analysis software that, at the time, increased the accuracy of the mass spectral data tenfold over the data acquired from the mass spectrometer, and incorporates algorithms that assist with the identification of peptides and proteins. The team published their first version of the method in *Nature Biotechnology* in 2008, just one year after Cox joined Mann's group[13]. Today, a proteomics study can involve the analysis of hundreds of thousands of peptides, Cox says, an impossible task without automated analysis. "Now researchers are freed up to spend more time thinking about biological processes and their implications," he says.

Jürgen Cox's software package takes the pain out of quantifying large collections of protein profiles.

Since their initial publication, Cox has become a group leader at MPI, and he and Mann and their colleagues have published additional methods that allow researchers to quantify proteins within samples without the use of labeled proteins; they continue to work on new ways to incorporate computational analysis and machine learning into proteomics research.

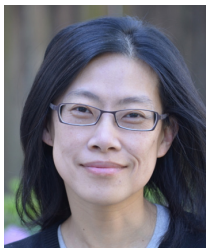
To help researchers make the most of MaxQuant, Cox and his colleagues offer a weeklong 'Summer School' at MPI each year. Last year, 250 researchers converged to learn the finer points of using these analysis tools and get expert assistance working on their own projects.

One endonuclease, many models

CRISPR-Cas9 has become the go-to research tool for making precise alterations to target genes. Not only is the targeting of Cas9 nuclease easy and simple to achieve through modification of the guide RNA, but the enzyme has also proven effective in the hands of many

investigators and in many different experimental models. Just how many different organisms are amenable to the tool became evident in early 2013 when papers started appearing in the literature—three of them in *Nature Biotechnology*[14–16].

CRISPR-Cas9 is derived from the adaptive immune systems of bacteria that contain within their genomes clustered, regularly interspaced, short palindromic repeats (CRISPRs). A series of discoveries about the function of CRISPRs and associated proteins, including Cas9, culminated in dramatic reports in the summer of 2012 that showed that the molecular phenomenon could be harnessed to site-specifically cleave double-stranded DNA *in vitro*[17,18].



Jing-Ruey Joanna Yeh says gene editing is not a niche enterprise anymore.

"Once that information was out people in this field knew that the game was on," says Jing-Ruey Joanna Yeh, a chemical biologist at Massachusetts General Hospital (Boston). A global race commenced to determine whether CRISPR-Cas9 would work in cultured cells, eukaryotic cells and whole organisms, not just in isolated DNA. "You can imagine there are a lot of things that could make it not work *in vivo*," says Yeh. "One never knows, without careful experiments, whether Cas9 can cleave chromosomal DNA wrapped around histones and packaged into chromatin in human or any other eukaryotic cells and induce genome editing," adds Jin-Soo Kim, director of the Center for Genome Engineering at the Institute for Basic Science, and professor at Seoul National University in South Korea.

By early 2013, Yeh, Kim and several other groups had independently—and within weeks of each other—reported their successes. "Papers came tumbling out," says Dana Carroll, at the University of Utah in Salt Lake City. "It was like, 'Bingo! It works,'" he says. In three papers that appeared in our March 2013 issue, Yeh's group used the tool to mutate the metabolic gene fumarate hydratase and nine other genes in zebrafish embryos; Kim and his team targeted the HIV co-receptor chemokine CC-motif receptor 5 (CCR5) in human cells; and Luciano Marraffini, a bacteriologist at The Rockefeller University in New York, and his team, used CRISPR-Cas9 to make multiple, simultaneous mutations in two types of bacteria, *Streptococcus pneumonia* and *Escherichia coli*.

Around the same time, papers appeared elsewhere in the literature, demonstrating the system in human and mouse cells[19–21]. *En masse*, the papers displayed the broad utility of CRISPR-Cas9 in diverse systems, says Kim.

What came next was a dramatic shift in life sciences research. Laboratories wanting to do genetic manipulations suddenly had a cheap and easy way to do it, enabling research that couldn't have been done before, says Carroll. And many labs that had been using other genome engineering tools or other methods for creating animal models switched to CRISPR. "Genome editing is not a unique niche for a few labs anymore," says Yeh.

Jin-Soo Kim obtained high frequencies of gene editing in human cells.

The upside of that, Yeh says, is that instead of spending so much time trying to perfect the tool, her laboratory can focus on applying the tool to answering more interesting biological questions. The downside, she says, is that there's a lot more competition.

Nucleotide resolution in nanopores

In 1996, a paper demonstrating that DNA can pass through protein pores in a membrane when a voltage is applied[22] prompted two scientists to sit up and take notice. The University of California, Santa Cruz (UCSC)'s Mark Akeson (then at the US National Institutes of Health; NIH) and University of Washington's (Seattle) Jens Gundlach ended up switching professional gears. Akeson left the NIH where he was studying G protein–coupled receptors to move to Santa Cruz, where David Deamer, one of the co-authors on the *PNAS* paper[22], had set up a sequencing unit; Gundlach, a gravitational physicist, along with one of the physics department's students, added the study

of nanopores to his research interests. More than 15 years later, the Gundlach and Akeson laboratories published back-to-back papers in *Nature Biotechnology*, reporting on a design that demonstrated the first viable nanopore sequencing device[23,24].

Early observations had indicated that the passage of a DNA strand through the membrane pore caused current changes that seemed related to sequence, but attempts to fine-tune base resolution were hampered because the molecule sped through the pore too quickly. In addition, early nanopores constructed with the protein α-hemolysin created a cavity that accommodated more than ten bases, which hindered recognizing individual bases. Several attempts were made to slow down the DNA by binding the template to various DNA processing enzymes, such as the Klenow fragment of *Escherichia coli* and others from the A family of polymerases, but the enzymes slipped off the DNA after only a few bases had passed through the pore.

Akeson's group at UCSC started working with an enzyme that binds stably to DNA, phi29 DNA polymerase, which reportedly replicates up to 70 kb of DNA after a single binding event[25]. But the question still unanswered was what would happen to the enzyme when bound to the DNA in the setting of an electric field.

Mark Akeson was ready to give up on nanopore sequencing until he found a way to hold the DNA onto the pore.

In a 2010 paper[26], Akeson and his group showed that the phi29 polymerase held the DNA on the pore some 40,000-fold longer than an A family polymerase. This was a turning point for many in the field, according to Akeson, who at the time was prepared to abandon trying to couple enzymes to the nanopore unless some dramatic advance came along. His 2012 *Nature Biotechnology* paper provided a further advance in control over the process by incorporating a blocking oligomer, which keeps the phi29 DNA polymerase from acting on the DNA in solution, and allows enzyme activation only when bound to the pore (in the presence of an electric field). The enzyme unzips, then copies the DNA, pushing it through the pore a base at a time, and feeding as many as 500 molecules through a single pore in single file order—essentially automating the process[23].

In a companion paper, Gundlach used a different nanopore, *Mycobacterium smegmatis* porin A (MspA), that had the right structure—a nucleotide-sized constriction—for resolving individual bases. A slight modification was required; the native protein repelled negatively charged DNA. Gundlach's group created a mutant that had no charge in the critical pore constriction[27]. Putting together these two elements—DNA bound with phi29 DNA polymerase and a nanopore comprising the engineered MspA—were key steps in the route to a nanopore sequencer; indeed, Gundlach's paper describes the sequencing of a 30-base oligomer[24].

A few years later, his group bested that by sequencing thousands of bases[28] and now, he says, there is no limit. "People have reported reads 100,000 bases long. It's not out of the question."

Cees Dekker, of Delft University's Kavli Institute (the Netherlands) says that these papers were important advances in nanopore-based DNA sequencing. "They were the first to show that one can get reproducible sequence information by analyzing the complex current signals upon traversing a string of DNA sequentially through the pore," he says.

Gundlach, who never really left physics, uses his nanopores for measuring the motion of single proteins. "We can measure the motion of DNA in enzymes as small as 40 picometers, which is totally unheard of in the field of biophysics. Most people do something like optical tweezers, which have the highest spatial resolution. Our spatial resolution using the nanopore is a factor of ten better[29]. Akeson, as well as Gundlach, have gone on to show the versatility of the nanopore; unlike conventional sequencing, nanopore sequencing can distinguish modified bases on the fly—no special modification of the template is required as it is for Sanger sequencing. Similarly, Gundlach's group also has shown that unnatural DNA bases can be distinguished.

Jens Gundlach never left his first field, physics, to develop nanopore sequencing.

Today, nanopore sequencing is beginning to make inroads in the commercial setting, pitting upstart Oxford Nanopore Technologies (where Akeson is a paid consultant and shareholder) against sequencing giant Illumina (San Diego), which has licensed University of Washington's technology (Gundlach is a paid consultant). Several nanopore technologies beyond protein pores are also in the works. Researchers are making progress with solid-state nanopores, made from graphene, for example, which are less fragile and would be easier to couple with silicon electronics, according to Dekker.

Seeing red

The richness of an image depends on the pigments with which one can paint. Just over a decade ago, multispectral imaging was limited by the color palette and number of available fluorescent proteins. Rational engineering of proteins based on green fluorescent protein (GFP), originally isolated from the jellyfish *Aequoria victoria*, had yielded blue, cyan, green and even yellow fluorescent labels. But filling in the red end of the visible spectrum had proven especially challenging, says Nathan Shaner, now a founding investigator at the Scintillon Institute, a nonprofit research institute in San Diego. That was until Shaner stepped into the primary role of building better red and orange proteins, culminating in a research publication in *Nature Biotechnology* in 2004 (ref. 30).

Nathan Shaner expanded the palette of fluorescent probes for imaging studies.

Isolated from a coral, the earliest red fluorescent protein, DsRed, was a tetramer, which made it impractical for following protein trafficking. At the University of California, San Diego, Roger Tsien's group had developed a monomeric version of the red protein, mRFP1, published in 2002, just before Shaner joined the group as a PhD student. But this new protein wasn't photostable.

Building a better red protein involved trial and error and a variety of educated guesses, says Shaner. "There was no crystal structure of the monomer at that time," he says. He and his colleagues randomized residues near the fluorescent chromophore to shift the absorption and emission spectra further in the red and introduced N-terminal residues that more closely resembled those in GFP to create more stable fusion proteins.

"The paper was really exciting because it introduced mCherry, which has become a mainstay of red fluorophores because of its brightness, monomeric character and photostability," says Jennifer Lippincott-Schwartz, at the National Institutes of Health. Another protein described in this paper, mOrange, was the brightest fluorescent labeling protein at that time.

"It was a fun time," Shaner recalls, developing a recognizable naming system for their rainbow of fluorescent colors. For a while the team threw around numerical systems or crayon colors, but eventually settled on fruit. (This paper also introduced other hues, such as mStrawberry and mTangerine.)

In addition to the widespread use of mCherry, TdTomato, a bright tandem dimer reported in that original study, still shows up regularly in research talks, Shaner says. And it's been gratifying to see the proteins show up in a host of imaging applications, including the brainbow mice. "It's fun to see something you put out into the world being used by a lot of people."

1. Baker, M. *Nat. Biotechnol.* **24**, 301–308 (2006).
2. Baker, M. & DeFrancesco, L. *Nat. Biotechnol.* **29**, 221–227 (2011).
3. Lister, R. *et al. Cell* **133**, 523–536 (2008).
4. Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L. & Wold, B. *Nat. Methods* **5**, 621–628 (2008).
5. Nagalakshmi, U. *et al. Science* **320**, 1344–1349 (2008).
6. Cloonan, N. *et al. Nat. Methods* **5**, 613–619 (2008).
7. Trapnell, C. *et al. Nat. Biotechnol.* **28**, 511–515 (2010).
8. Grabherr, M.G. *et al. Nat. Biotechnol.* **29**, 644–652 (2011).
9. Trapnell, C. *et al. Nat. Biotechnol.* **31**, 46–53 (2013).
10. Hass, B. *et al. Nat. Protocols* **8**, 1494–1512 (2013).
11. Langille, M.G.I. *et al. Nat. Biotechnol.* **31**, 814–821 (2013).
12. Iyer, S.M. *et al. Nat. Biotechnol.* **32**, 274–278 (2014).
13. Cox, J. & Mann, M. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
14. Hwang, W.Y. *et al. Nat. Biotechnol.* **31**, 227–229 (2013).
15. Cho, S.W., Kim, S., Kim, J.M. & Kim, J.S. *Nat. Biotechnol.* **31**, 230–232 (2013).
16. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L.A. *Nat. Biotechnol.* **31**, 233–239 (2013).
17. Jinek, M. *et al. Science* **337**, 816–821 (2012).
18. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. *Proc. Natl. Acad. Sci. USA* **109**, E2579–E2586 (2012).
19. Cong, L. *et al. Science* **339**, 819–823 (2013).
20. Mali, P. *et al. Science* **339**, 823–826 (2013).
21. Jinek, M. *et al. eLife* **2**, e00471 (2013).
22. Kasianowicz, J.J., Brandin, E., Branton, D. & Deamer, D.W. *Proc. Natl. Acad. Sci. USA* **93**, 13770–13773 (1996).
23. Cherf, G.M. *et al. Nat. Biotechnol.* **30**, 344–348 (2012).
24. Manrao, E.A. *et al. Nat. Biotechnol.* **30**, 349–353 (2012).
25. Blanco, L. *et al. J. Biol. Chem.* **264**, 8935–8940 (1989).
26. Lieberman, K.R. *et al. J. Am. Chem. Soc.* **132**, 17961–17972 (2010).
27. Butler, T.Z., Pavlenok, M., Derrington, I.M., Niederweis, M. & Gundlach, J.H. *Proc. Natl. Acad. Sci. USA* **105**, 20647–20652 (2008).
28. Laszlo, A.H. *et al. Nat. Biotechnol.* **32**, 829–833 (2014).
29. Derrington, I.M. *et al. Nat. Biotechnol.* **33**, 1073–1075 (2015).
30. Shaner, N.C. *et al. Nat. Biotechnol.* **22**, 1567–1572 (2004).
31. Lockhart, D.J. *et al. Nat. Biotechnol.* **14**, 1675–1680 (1996).
32. Tyagi, S. & Kramer, F.R. *Nat. Biotechnol.* **14**, 303–308 (1996).
33. Gygi, S.P. *et al. Nat. Biotechnol.* **17**, 994–999 (1999).
34. Zheng, G., Patolsky, F., Cui, Y., Wang, W.U. & Lieber, C.M. *Nat. Biotechnol.* **23**, 1294–1301 (2005).
35. Lu, P., Vogel, C., Wang, R., Yao, X. & Marcotte, E.M. *Nat. Biotechnol.* **25**, 117–124 (2007).
36. Ball, M.P. *et al. Nat. Biotechnol.* **27**, 361–368 (2009).
37. Deng, J. *et al. Nat. Biotechnol.* **27**, 353–360 (2009).

Corrected after print 31 March 2016.

Erratum: 20 years of *Nature Biotechnology* research tools

Anna Azvolinsky, Laura DeFrancesco, Emily Waltz & Sarah Webb
***Nat. Biotechnol.* 34, 256–261; published online 10 March 2016; corrected after print 31 March 2016**

In the version of this article initially published, on p. 257, Steven Salzberg's name was misspelled as "Stephen." The error has been corrected in the HTML and PDF versions of the article.

Erratum: 20 years of *Nature Biotechnology* biomedical research

Anna Azvolinsky, Charles Schmidt, Emily Waltz & Sarah Webb
***Nat. Biotechnol.* 34, 262–266; published online 10 March 2016; corrected after print 31 March 2016**

In the version of this article initially published, the leadership of the team developing ADCs was incorrectly attributed, for which Nature Biotechnology takes full responsibility. To correct this, on p.265, paragraph 3, "Shen's team at Genentech" now reads "Jagath R. Junutula (then at Genentech, now at Cellerant Therapeutics in San Carlos, CA, USA) and colleagues"; on p.266, first line, "Since then, Shen's group" now reads, "Since then, Genentech"; and the photo of Shen with the caption, "Ben-Quan Shen's group at Genentech found a way to reduce the complexity of antibody-drug conjugates" has been replaced with a photo of Junutula with the caption, "Jagath R. Junutula's group highlighted the important role of site-specific conjugation in ADC development." In addition, Kadcyla was misspelled as Kadcycla, and Adcetris as Adcentris. The errors have been corrected in the HTML and PDF versions of the article.

Erratum: When biotech goes bad

John Hodgson
***Nat. Biotechnol.* 34, 284–291 (2016); published online 10 March 2016; corrected after print 31 March 2016**

In the version of this article initially published, on p.286, in the timeline, far right, "2008 Intercept" should have read "2008 InterMune." The error has been corrected in the HTML and PDF versions of the article.

Erratum: Community crystal gazing

Anu Acharya, Kate Bingham, Jay Bradner, Wylie Burke, R Alta Charo, Joel Cherry, André Choulika, Tony Coles, Robert Cook-Deegan, Stanley T Crook, Emilia Díaz, Brent Erickson, L Val Giddings, Sebastian Eriksson Giwa, James C Greenwood, Vishal Gulati, Sam Hall, John Harris, Jamie Heywood, Colin Hill, Jeremy Levin, Adina Mangubat, John Maraganore, Giovanni Mariggi, Barbara J Mazur, Amy L McGuire, Nathalie Moll, Jonathan Moreno, Gail Naughton, Lita Nelsen, Jane Osbourn, Daniel Perez, John Reed, Eric Schmidt, Vicki Seyfert-Margolis, Paul Stoffels, Jørgen Thorball, Tara O'Toole, Indrek Vainu, Sander van Deventer, Elias Zerhouni & Daphne Zohar
***Nat. Biotechnol.* 34, 276–283 (2016); published online 10 March 2016; corrected after print 31 March 2016**

In the version of this article initially published, Stanley Crooke's name was misspelled as "Crook" in the author list. The error has been corrected in the HTML and PDF versions of the article.

Corrigendum: Combinatorial hydrogel library enables identification of materials that mitigate the foreign body response in primates

Arturo J Vegas, Omid Veiseh, Joshua C Doloff, Minglin Ma, Hok Hei Tam, Kaitlin Bratlie, Jie Li, Andrew R Bader, Erin Langan, Karsten Olejnik, Patrick Fenton, Jeon Woong Kang, Jennifer Hollister-Locke, Matthew A Bochenek, Alan Chiu, Sean Siebert, Katherine Tang, Siddharth Jhunjhunwala, Stephanie Aresta-Dasilva, Nimit Dholakia, Raj Thakrar, Thema Vietti, Michael Chen, Josh Cohen, Karolina Siniakowicz, Meirigeng Qi, James McGarrigle, Stephen Lyle, David M Harlan, Dale L Greiner, Jose Oberholzer, Gordon C Weir, Robert Langer & Daniel G Anderson
***Nat. Biotechnol.* 34, 345–352 (2016); published online 29 January 2016; corrected after print 18 April 2016**

In the version of this article initially published, one author, Adam C. Graham, his affiliation, and his contribution were omitted. In addition, two acknowledgments, to W. Salmon and J. Wyckoff, were omitted. The errors have been corrected in the HTML and PDF versions of the article.