# Five more years of *Nature Biotechnology* research

Monya Baker & Laura DeFrancesco

**Authors of the past five years' most highly cited research articles discuss their work and new directions in their respective areas.**



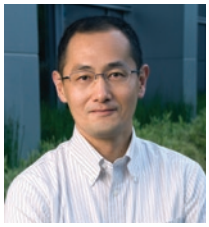
Five years ago, one of the ways we celebrated our tenth anniversary as *Nature Biotechnology* was by looking back at the most highly cited papers of the previous decade. As we enter our fifteenth year, we take another look back, this time just five years. During this brief interval, assumptions about how stem cells can be generated and differentiated were revised, new technologies emerged in protein analysis and sequencing allowing 'omics approaches to move from species to individuals to differences between individuals, and computer models got better at predicting cellular behavior according to our knowledge of biological pathways. Because of space constraints, we do not cover all the biotech advances from the past five years; instead, the vignettes below provide a selection of the most important advances published in our pages with a nod to the implications and further applications of the work.

## Safer iPS cells


One of the most important discoveries since the creation of induced pluripotent stem (iPS) cells began as a negative control experiment. To make iPS cells, Shinya Yamanaka of Kyoto University showed that inserting just four transcription factors into cultured mouse fibroblasts could make them behave like embryonic stem cells[1]. All four proteins were considered essential, but they also presented risks for any potential therapeutic use of reprogrammed cells. Indeed, one of them, *c-Myc*, is a well-known oncogene.

*Monya Baker is Technology Editor for Nature and Nature Methods. Laura DeFrancesco is Senior Editor at Nature Biotechnology.*

In experiments aimed at understanding the role of Myc in reprogramming, Yamanaka's laboratory set up experiments to study c-Myc compared with other members of the gene family and included a negative control without Myc. "To our surprise, we obtained iPS cells, even without Myc, albeit with a very low efficiency," says Yamanaka.



Shinya Yamanaka says the Myc genes may hold the key to other aspects of reprogramming.

Yamanaka went on to make Myc-free iPS cells from both mouse and human cells[2]. To determine whether the presence of c-Myc increases the chance that the iPS cells will form tumors, Yamanaka's group mixed mouse iPS cells into early-stage mouse embryos, which grew into chimeric mice with many tissues derived from the iPS cells. Six of 37 mice made from Myc-containing iPS cells died from tumors, whereas none of the 26 mice generated from Myc-free iPS cells did.

Since this work was done, researchers have been looking hard for ways to make iPS cells without inserting any transgenes at all; for example, by using combinations of small molecules, integration-free DNA vectors, synthetic RNA encoding the transcription factors and even modified versions of the proteins themselves. But, as Yamanaka warns, merely eliminating Myc or integrating factors does not guarantee the safety of iPS cells. Some iPS cell–derived tumors could be caused by reactivating the gene encoding Myc, others by residual undifferentiated cells that resist cues for differentiation, a property that correlates with the tissue from which the iPS cells were originally derived[3]. And, according to Yamanaka, there may be other causes, so the safety of iPS -cells needs to be rigorously assessed before they can be used in the clinic.

Meanwhile, Yamanaka says the Myc gene family may hold the key to other aspects of reprogramming. Last year, his group found that mutant versions of c-Myc, as well as the related protein L-Myc, could all promote reprogramming to iPS cells more efficiently than c-Myc, even though they have little transformation activity on their own[4]. Moreover, L-Myc does not promote tumor formation in mice.

These results show that Myc family genes can promote reprogramming independent of their ability to make tumors. How this happens is still unclear. There are several possible mechanisms, says Yamanaka. Myc could affect chromatin architecture and function synergistically with other transcription factors; it could also regulate genes that in turn regulate cell differentiation. By delving further into the molecular mechanisms of reprogramming and identifying more factors, generating safer iPS cells may become faster and more predictable. And perhaps the next discovery will also come from negative controls with surprising results.

## Why some tumors don't starve


After a bit of evangelism and enough clinical data, the idea seemed obvious: drugs that deprive tumors of their blood supply should shrink them away to nothing. In practice, though, some tumors stayed put. Though resistance in tumors is common against toxic drugs that attack cancer cells directly, it was not expected to develop against angiogenesis inhibitors, which target presumably normal vascular epithelial cells rather than fast-dividing, quickly mutating cancerous ones.

Vascular endothelial growth factor (VEGF), the signaling molecule that summons blood

Napoleone Ferrara was given the freedom to pursue his interest in tumor angiogenesis.

vessels, was discovered in 1989 by Napoleone Ferrara, a cancer biologist at Genentech (South San Francisco, CA, USA). The discovery was actually a side project; Genentech's philosophy was to give scientists freedom to pursue their own interests part-time. This paid off, as Avastin (bevacizumab), a humanized monoclonal antibody drug directed against VEGF that blocks this signaling, was approved in 2004 for metastatic colon cancer. When resistance to Avastin was observed, Ferrara decided to hunt down the mechanism responsible. Because Avastin stops cancer growth only *in vivo*, not *in vitro*, he reasoned biology outside the tumors had to be involved.

Ferrara and his colleagues identified cell lines that, when grown in mice, produced tumors that were either sensitive or refractory to the anti-VEGF antibody. Thus began a long series of animal experiments. His team quickly ruled out that a specific immune attack was involved: anti-VEGF antibodies inhibited sensitive tumors in immunodeficient mice, and refractory tumors behaved the same in immunocompetent and immunodeficient mice.

Ferrara suspected that studying inflammation might yield the answer to resistance; many reports at the time showed that inflammation played a role in cancer, but no one had looked at whether it might play a role in response to anti-VEGF treatments. In fact, regardless of treatment with the anti-VEGF antibody, refractory tumors contained many more bone marrow mononuclear cells (BMMNCs) and more vasculature than sensitive tumors, even though the BMMNCs were not contributing to blood vessels, at least not directly.

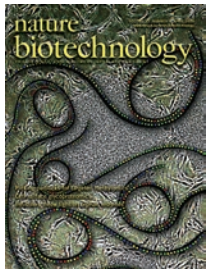
To figure out what role the BMMNCs were playing, Ferrara's team exposed the cells to extracts from sensitive and refractory tumors. After exposure to extracts from resistant tumors, BMMNCs promoted growth of sensitive tumors. Somehow, the refractory tumors 'instructed' BMMNCs to stimulate tumor growth.

To unravel the mechanisms, Ferrara and his collaborators first had to figure out what type of BMMNCs were responsible. This work eventually fingered a subset of myeloid cells. The myeloid cells were promoting blood vessel growth by secreting factors in addition to VEGF. Suppressing myeloid cells made Avastin-refractory tumors more sensitive[5].

Identifying the specific factors required additional work. In brief, either stroma or tumor cells themselves can produce colony-stimulating factors (in this case, granulocyte-macrophage colony-stimulating factor and granulocyte colony-stimulating factor) that mobilize myeloid cells, which in turn promote angiogenesis through a protein called BV8 (refs. 6,7). Although these experiments were done in mice, evidence is growing, says Ferrara, that a similar process also occurs in humans, though it is still unclear which subset of myeloid cells mediates this effect. (Right now, most scientists are betting on neutrophils.) In addition, several other reasons have been proposed for why tumors don't respond to anti-angiogenesis drugs, such as poor delivery of the drug to the tumor or the presence of redundant signaling pathways promoting angiogenesis.

Although the mechanism of resistance is complicated, Ferrara says the results of anti-VEGF treatments has still come as a shock. In the early 1990s, those targeting tumor angiogenesis assumed that to see any efficacy, they would need to block many factors, until it became clear that knocking out VEGF alone could stop the blood supply at least in some cases. "What surprised me was that VEGF could be such an important molecule."

## Next-generation sequencing captures methylation variation




People are thinking less and less about 'the' human genome and 'the' human epigenome, says George Church of Harvard University (Cambridge, MA, USA). Instead, researchers are trying to home in on variation, comparing genomes and epigenomes across samples. In 2009, two papers in *Nature Biotechnology* described efficient ways to do so.

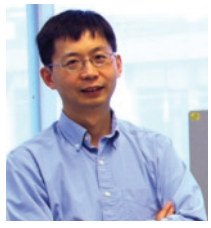
One of the techniques described by Church and his colleagues relies on restriction enzymes that preferentially cut unmethylated sequences. Intense sequencing around the restriction sites can compare nearly 1.5 million restriction sites between samples with considerably less effort than whole-genome approaches require. A less comprehensive but more reproducible method uses 10,000 'padlock probes', a clever technique for allowing many different sequences of DNA to be amplified in the same reaction vessel using a common set of primers[8].

To use padlock probes to detect methylation, DNA is first treated with bisulfite, which chemically converts unmethylated cytosine so that it registers as thymine. It was a risky project, says Kun Zhang, who completed his postdoc with Church and is now at the University of California, San Diego. "Using padlock probes to detect methylation was really a bet. Once DNA is converted, the majority of the genome is made of just three members of the genetic code, so the genome complexity goes down dramatically."

George Church says that commonly held but false beliefs among scientists impede technology development.

In a separate study, Zhang designed some 30,000 probes to assess methylation at 66,000 sites in the genome and identified methylation differences between pluripotent and specialized cells[9].

The key to success was being able to make many padlock probes[9]. For this, Church and Zhang hit upon the idea of using microarrays as miniature DNA synthesizers: rather than using the array to detect DNA in a sample, the array is used to make probes, which are then clipped off. It wasn't easy, recalls Church. "Murphy's law was in full force. Almost every aspect would fail. Some chips had a 1% yield or a nonrandom yield, so you'd get some element overrepresented by a couple factors of ten." Zhang began a collaboration with Agilent Technologies (Santa Clara, CA, USA), and the numbers of probes grew from 10,000 to 20,000 to over 55,000, recalls Zhang. "Once we had the capability of making tens of thousands of probes relatively easily, we could start to apply probes to a real study."



Kun Zhang says that using padlock probes to detect methylation was really a bet.

Zhang is currently working on ways to increase throughput still further, and points to several improvements: the original protocol required ten purification steps, from introducing the probes to sequencing the new libraries, says Zhang. Now that is down to three steps.

And the padlock probes are much improved as well. "Initially we didn't understand all the parameters, so what we did was just brute force," says Zhang. However, after measuring the performance of probes and feeding that data into a machine-learning algorithm, probes with much better performance

are now being designed, he says. In the first papers, only ~50% of the reads were useful. Now, he says, ~80% are useful, a rate much closer to regular genome sequencing. The coverage of the genome is becoming both more complete and more uniform, he says.

The idea of using padlock probes comprehensively was once considered unthinkable, if only for the difficulty of the DNA synthesis, says Church. It's just one example where commonly held but false beliefs among scientists impede technology development, he says. "Sometimes they'll say something is impossible, but if you drill down they say that it's really expensive, and if you drill further, they say it's really expensive now."

## Nanoprobes, over and out, but not finished



Getting an imaging probe out of the body can be just as important as putting it in. That's especially true for quantum dots, which fluoresce brightly but are made from toxic elements. Thus, anyone interested in using dots for clinical applications needs to understand how they move through the body. Researchers led by John Frangioni at Beth Israel Deaconess Medical Center in Boston and Moungi Bawendi at the Massachusetts Institute of Technology in Cambridge described how they tracked dots with a wide range of physical parameters and showed that, at least in mice, quantum dots can be cleared through the kidney and into urine, provided that the dots are fewer than 6 nm across, about half the size of an antibody[10].

But quantum dots pose so many problems for human applications that Frangioni has decided to focus instead on alternative imaging agents. Even the smallest dots don't clear as completely or quickly as he would like. The color of nanodots depends on both size and composition, and to get the most useful colors for imaging, very small quantum dots must be made of some of the most toxic elements. Furthermore, staying below 6 nm is difficult once dots are coated to prevent leaching and functionalized with molecules to target them to particular tissues.

John Frangioni says, "We hope our work will make others think about what's potentially clinically viable and what's a dead-end."

Shuming Nie thinks that gold nanoparticles will have multiple clinical uses.

Despite what enthusiasts say, quantum dots are not particularly bright once they are in the body, says Frangioni. It is true that quantum dots outside the body can be stimulated with a broad bandwidth of light frequencies, but in scattering tissue, absorption is usually confined to a narrow, redder band. And on a per-volume basis, small molecules are brighter than quantum dots, says Frangioni. Thus, after spending almost a decade working on quantum dots for medical imaging[11], he has decided to devote his efforts to small-molecule dyes instead. "The periodic table is just not in our favor," he says. "We hope our work will make others think about what's potentially clinically viable and what's a dead-end."

Other researchers are continuing work on nanotech imaging probes for clinical use, however, and one of these is Shuming Nie, who has appointments at both Emory University and the Georgia Institute of Technology in Atlanta. In 2008, he reported a gold-based nanoparticle that could be conjugated to an antibody and used to detect cancer[12]. Before this paper, the brightest tools used in cancer imaging were quantum dots. Not only is gold already used in humans, Nie was able to produce gold nanoparticles 200 times brighter than any previously reported.

The brightness of the nanoparticles is a function of the gold itself as well as small-molecule dyes adsorbed to it, explains Nie. One potential application of these particles is demarcating the edges of tumors, allowing surgeons to be sure they remove all the cancerous tissues. Last year, his laboratory described a proof-of-principle experiment showing that a handheld device termed a SpectroPen could be used along with these nanoparticles to visualize the borders of tumors implanted in mice[13].

Nie thinks the particles could be used in other ways as well. The surface-enhancing Raman scattering dyes adsorbed to the particles come in a variety of colors, and so more than one kind of molecule could be imaged simultaneously within the same animal. It's even possible, he says, that rare cells within an animal could be tagged with these nanoparticles and imaged over time. There are some problems that will have to be assessed or mitigated before clinical trials, says Nie. For instance, the nanoparticles accumulate in the liver and other organs. But Nie believes such issues can be overcome eventually. "I agree that for systemic applications quantum dots are not the way to go. These [gold] particles are really different."

## Picking the right kinase


Many commercially successful kinase inhibitors have off-target effects, but until recently few researchers realized how extensive these could be. A few years ago, Ambit Biosciences (San Diego) introduced the idea of a 'selectivity score' based on the company's technology for profiling hundreds of kinases against thousands of compounds[14,15].

Using data collected by the company's KinomeScan profiling technology, these analyses produce kinase trees covered with dots, which show each compound's selectivity and potency across the universe of kinases, or the kinome. These images are now iconic, but initially, the data were visualized quite differently, says author Patrick Zarrinkar. For example, data for multiple compounds would be shown on a single tree image. Then a collaborator requested an image showing data for just a single compound. "As soon as I saw that, it was clear that this would be the way we should display the profiling results," he recalls.

Patrick Zarrinkar says, "It was so many small breakthroughs, we'd just put our heads down and continue to work."

Pharmaceutical teams were surprised to learn that molecules they had been working with for years were less selective than they'd assumed. "They said 'we thought we understood this compound, but it does all these other things,'" recalls Zarrinkar. "Before these papers, it was normal for people to take an inhibitor and test against five or six kinases and assume it was selective," says Zarrinkar, now at Prognosys Biosciences (La Jolla, CA, USA). "You can't do that today."

Ambit's and others' data showing that clinical kinase inhibitors hit multiple targets startled researchers into seeking broader profiling of potential drug candidates. This has

changed how people approach drug development. Now, says Zarrinkar, pharmaceutical companies are not just looking for molecules that hit a predefined target, but are asking which interesting targets can be hit by their molecules. "You can make decisions not just based on the biology of the target but also based on chemistry—which targets are hit by available compounds," says Zarrinkar.

This is exactly what Ambit did when selecting its own clinical candidate, now in phase 2 trials for acute myelogenous leukemia. After screening its compound library against a panel of kinases, the company picked a disease target, FMS-like tyrosine kinase 3 (FLT3), for which their library had the best hit. Some advisors discouraged them, citing evidence from FLT3 inhibitors already in the clinic that suggested it wasn't a good target. After comparing the profiles between the compounds that were in the clinic and the hit Ambit had identified, the company decided its compound's selectivity and potency profile was more promising and decided to go forward. Subsequently, it sold the KinomeScan technology to DiscoveRx (Fremont, CA, USA) and is focusing on clinical development of this and other drugs.

Ironically, the decision to develop the kinase profiling assay in the first place may have benefited from having less information, says Zarrinkar. Ambit had been developing phage-display technology for a different purpose and realized it could be used to quickly build extensive panels of kinases for assays. Fortunately, company researchers were unaware how difficult some kinases could be to produce, says Zarrinkar. "Everyone told us we were crazy, but we didn't know kinases, so we thought we could figure it out." They did, though with considerably more effort than they had expected. Progress was slow and incremental, says Zarrinkar. "It was so many small breakthroughs, we'd just put our heads down and continue to work." Eventually, team members including David Lockhart and Daniel Treiber realized they had to be very careful about which section of the kinase was displayed and how. In fact, many assays in the current version of KinomeScan rely on proteins produced in mammalian cells rather than on phage.

Each profiling assay measures the binding equilibrium between a kinase, a test compound, and a known, small-molecule ligand. A biotinylated ligand is immobilized on streptavidin-coated beads and introduced to a free test compound and a kinase. Kinases are tagged with a stretch of DNA, and so quantitative PCR can assess the number of bound kinases sensitively and over a broad dynamic range. The binding properties of kinases to the immobilized ligands are used to assess the test compounds' affinity. These assays can be easily multiplexed and allow tens of thousands of compounds to be assessed against hundreds of kinases quickly.

Zarrinkar says that this profiling technology represents one way in which the human genome project has directly benefited drug discovery. Rather than enabling the discovery of novel targets, the genome is enabling the discovery of novel compounds. Systematic development of the kinase panel was made possible by the enumeration of the kinome by other scientists, which in turn required their having the human genome sequence[16]. The profiling technology then enabled a novel approach to kinase inhibitor discovery[17], says Zarrinkar. "There's a straight line from the human genome to the kinome to KinomeScan technology to novel drugs."

## Quality control for microarrays


In the early twenty-first century, after a few years of unbridled excitement, doubts began to be raised about the validity of microarray results. Different laboratories, often using different platforms, were getting different results. Still, no one could deny that the technology offered powerful information: a read on what genes were being turned on, off, up and down inside cells.

Even as the community was questioning the technology, the US Food and Drug Administration (FDA) was investing considerable effort embracing pharmacogenomics and thinking about how to incorporate genomic data submissions associated with drug applications. Leming Shi, a computational chemist at the FDA, believed something had to be done.

Shi organized a large-scale, quality control effort to solve these problems. He convinced Applied Biosystems (Foster City, CA, USA) to supply quantitative PCR data—the gold standard for gene expression measurement—to compare multiparallel data from DNA microarrays. Several microarray manufacturers as well as academic scientists joined the effort. Eventually, Shi attracted 137 participants from 51 organizations; people with competing interests and different scientific views were able to work together under the leadership of a federal agency. In addition to supplying microarrays and reagents, manufacturers even provided information about proprietary probe sequences. Collectively, laboratories analyzed two common reference RNA samples on seven microarray platforms, plus three other technologies for independently assessing expression. Ultimately, the project involved over 1,000 microarrays and represented a cumulative investment of around $2 million, says Shi.

Whereas the project covered many experimental and analytical facets of quality control, how to select informative genes became the most controversial. From the beginning, Shi says, the problem was figuring out which subset of genes provides biological insights. Microarrays simultaneously monitor the activity of tens of thousands of genes in one experiment, and then individual genes must be compared between different sets of samples. Researchers had developed different ways to decide which genes to focus on. One leading approach picked genes based on a metric called the *t*-statistic, a measure derived from variation in gene expression measurements. Unfortunately this varied substantially from experiment to experiment, so Shi didn't think this choice made sense. "People were dominated by pure statistical consideration without thinking of what the technology was trying to tell us," he says.

He asked participating scientists to analyze data in whatever way seemed best. "The only data that was reproducible was if you looked at the magnitude of the difference between the two conditions and then ranked all the genes on the chip based on the fold-change," Shi says. Using those criteria, data were much more reproducible across different laboratories and across different platforms. The group, which named itself the MicroArray Quality Control Consortium (MAQC), published its results in a research article[18] with five companion articles in 2006.

FDA computational chemist Leming Shi organized a large-scale quality effort to address the problems with microarray experiments.

But Shi hadn't appreciated how invested many scientists were in the *t*-statistic. When presenting an initial analysis of his results, one professor interrupted his talk, demanding that he recheck his results; Shi was told to restrict the MAQC studies to analyzing the amount of technological noise and to steer clear of evaluating statistical methods. Little by little, however, the statisticians within the MAQC were won over. A companion MAQC paper by Shi describing the rationale for ranking genes by fold-change was rejected in 2006 as unsound; two years later, another set of reviewers rejected the same paper as obvious.

Meanwhile, the MAQC has moved on. A recent publication in *Nature Biotechnology*

evaluated different approaches of using microarray data to create and validate predictive models[19]—so-called genomic classifiers—and is providing techniques and guidelines to avoid over-fitting data. The current project, says Shi, is evaluating data from next-generation sequencing. "We are seeing similar things where the performances of various platforms and data analysis approaches have not been adequately vetted before their widespread applications," he says, but it's "even more challenging for next-generation sequencing because the data are more complex and prodigious, and the applications are more diverse."

## Modeling human metabolism


Computer programs that model microbial metabolic networks have been around for nearly two decades, but human cells are trickier. The models for microbes constrain themselves to intuitive 'metabolic objectives': taking nutrients from the environment and using them to grow. Building such models requires detailed data about enzyme and metabolite concentrations over time, but such data are impractical to collect for large-scale modeling of human tissues, whose metabolic objectives are also not so easily defined. A bacterial cell that grows as fast as possible is probably an evolutionary success; a human cell that does the same is probably cancerous.

Tomer Shlomi says, "[Modeling] lets you use something that is straightforward to measure and predict something that is hard to measure on a large scale."

Computational biologists Eytan Ruppin and Tomer Shlomi at Tel Aviv University decided that, because they couldn't get the data they thought they needed, they should find a way to apply data they had. Starting from a model for a generic human cell developed by the Bernhard Palsson lab at the University of California, San Diego[20], Ruppin and Shlomi used microarray and proteomic data to predict metabolic activity specific for ten tissue types[21]. "It lets you use something that is straightforward to measure and predict something that is hard to measure on a large scale," explains Shlomi, now at the Technion Israel Institute of Technology in Haifa.

Eytan Ruppin says, "Building models is a matter of fine art, experience and scientific taste."

The model correctly predicts metabolic fluxes in yeast cells. In human tissues, it indicates post-transcriptional regulation for about a fifth of tissue-specific genes. The model also accurately categorizes tissue-specific activity of dozens of known disease genes, and it is already being applied in other settings; researchers who want to look at the fluxes and post-transcriptional regulation predicted by their own data can use freely available software called iMAT (Integrative Metabolic Analysis Tool)[22].

Meanwhile, Shlomi and Ruppin have extended their work to make more precise models of liver and cancer metabolism. The latter model predicted certain proteins that are particularly important in a hereditary form of kidney cancer and compared the list against known targets for anticancer drugs. So far, one potential new target has been experimentally validated by collaborator Eyal Gottlieb, from the Beatson Institute for Cancer Research (Glasgow, UK), says Shlomi; back-to-back papers describing the model and the target should be published later this year.

The researchers hope to make these metabolic models more sophisticated. Right now, the *in silico* networks do not include proteins involved in genetic regulation or cell signaling. And the model considers expression only qualitatively: genes can be expressed at high levels, low levels or not at all. Being able to include additional regulatory and signaling constraints and to use continuous, quantitative data for expression could improve the accuracy of predictions, says Shlomi.

Numerous laboratories in addition to Ruppin's and Shlomi's have been able to capture sophisticated aspects of brain, kidney, liver and other organs. Building models, says Ruppin, "is a matter of fine art, experience and scientific taste," combined with the drive to validate a model against relevant data. "In the kind of work that we do," says Ruppin, "you don't start from first principles and rigorously prove everything you do. You have to make some choices that may depend on your gut feeling of what might better reflect biological reality. Then you have the responsibility to check the robustness of the decisions." No matter how sophisticated, however, models can only give an approximate description of the underlying biology. "What we have," Ruppin says, "are drafts that should be improved upon." If those drafts point to new pathways and targets, they have already demonstrated their value.

## Wired sensors


In 2005, researchers led by Charles Lieber at Harvard University reported that nanowires coated with antibodies could detect five cancer biomarkers in serum with femtomolar sensitivity[23]. Rather than requiring laborious DNA amplification or instrument-intensive fluorescent labels, the wires detected the biomarkers through changes in conductance caused when electrically charged proteins bound to the silicon wire.

Charles Lieber says, "The big challenge is not to improve [technology], but to take what already has unique attributes and make a product that people can use."

More than a biomarker's presence could be detected. When functionalized with the appropriate DNA sequence, nanowires could also be used to monitor telomerase binding to and elongating the wire-bound primer. The nanowire system could even be used to screen for telomerase inhibitors. And the nanowires were poised for extreme multiplexing. Lieber predicted at the time that a basic nanowire sensor chip could potentially contain 200 or so individually functionalized wires, each its own nano-assay.

Since that paper, researchers have built other nanotech detection devices based on graphene, nanotubes and other materials. Proof-of-principle experiments have been carried out for multiplexed and real-time detection for many chemical species, but they are not in mainstream use.

Moving the technology from stunning proof of principle to useful tool requires the right combination of knowledge about semiconductors, surface chemistry and biology. An example within Lieber's own laboratory shows how seemingly trivial oversights can mean big delays. After demonstrating that the nanowire sensors worked with biotin and streptavidin[24], Lieber wanted a more biologically relevant example, such as detecting cancer biomarkers in serum. An engineering postdoc

worked on the problem for two years without making any progress, says Lieber. "Then this guy [Fernando] Patolsky came into the lab, and in three weeks everything was solved." The commercial antibodies that Lieber's laboratory had purchased were stabilized in bovine serum albumin, and the postdocs had unwittingly been attaching the albumin rather than the antibody when fabricating the nanowires. In fact, Lieber is hopeful that co-author Patolsky might be able to commercialize the nanosensor technology at a new startup, now that another company, Nanosys (Palo Alto, CA, USA), which owns the relevant intellectual property, has indicated some willingness to sublicense.

Using the nanowire technology to make a single chip that detects a variety of biomarkers in real time is doable, but it's not something that people used to working with conventional techniques readily consider. "People are used to their ingrained technologies," Lieber says. "Maybe the big challenge for the field is not to improve it, but to take what already has unique attributes and make a product that people more generally can use."

Lieber is taking that idea with him into his current project, kinked nanowires capable of monitoring chemistry inside a living cell[25]. "We need to make biologists something that they know how to use," he says. Though the kinked nanowires have been put onto chips, neurobiologists are more familiar with patch clamps, so Lieber is looking into putting the nanowires on something resembling these tools.

At this stage, the key to making nanotechnology detection devices widespread is not always in pushing the science forward, says Lieber; it's often in improving usability, a task that is outside the scope of an academic laboratory. "The limitation and also frustration as a scientist is the commercialization," he says.

## Eying islets


On his first day at work at a fledgling biotech company in San Diego, Ed Baetge found rather less than he'd expected. As a newcomer to Cythera (which later became Novocell and is now ViaCyte), his task was to turn embryonic stem cells into insulin-secreting beta cells, but he didn't even have the starting material.

About a year later, in 2002, colleague Alan Agulnick produced the first line. "It was just a tiny little colony growing up on a bed of fibroblast feeders, and we had to do needle passage, where you would dissect away the undifferentiated material, all done morphologically on a dissecting scope." Other lines followed, and Baetge and his colleagues began puzzling out how to move cells from a pluripotent state to a pancreatic one.

Ed Baetge says, "[Islet cells are] probably the most difficult biological product you could ever consider."

The first step was to make definitive endoderm, one of the three germ layers of nonreproductive tissue, and the one that gives rise to pancreatic tissue. The goal was to mimic normal development, says Baetge. They began lowering the concentration of serum in the culture media and also adding the growth factor activin. The key to success was constant monitoring of gene expression in response to daily changes of media. "The culture had to be examined not on a daily basis but on an hourly basis," Baetge recalls.

The next steps involved applying various morphogens and growth factors, which finally resulted in the production of endocrine cells and the 2006 *Nature Biotechnology* paper[26]. The step that mattered, though, was the previous one[27]. "If we hadn't gotten good, clean, definitive endoderm, we never would have gotten there."

Still, endocrine tissue produced was a long way from functional pancreatic cells. For a long time, says Baetge, the cells produced little insulin in response to glucose and didn't express pancreatic markers stably. "We used to call them schmendocrine cells, a schmucky kind of endocrine cell," recalls Baetge.

Extrapolating from published research, Baetge speculated that progenitors needed some signals from mesoderm tissues to form functional pancreatic cells, but the signals didn't seem to be present in the *in vitro* culture system, so he and colleagues came to the surprising idea of transplanting the cells into mice. The thinking, he recalls, was "why not take the cells and put them *in vivo* and see if some of the signals they were missing would be delivered?"

After four weeks and no sign of insulin, one principal investigator was ready to stop the experiment, but Baetge urged them to wait. He had to urge again at eight weeks and at ten weeks, and then—finally— in the twelfth week, they found insulin being secreted in response to glucose[28]. The length of time is not so surprising, Baetge says, considering that it takes at least 100 days after conception for a developing embryo to develop bona fide islet cells. Now, researchers at ViaCyte are working on ways to make larger volumes of cells and shield them from the immune system. The progress has been impressive, says Baetge, now head of the Nestlé Institute of Health Sciences in Lausanne, Switzerland, but there is still a long way to go. "It's probably one of the most challenging biological products you could ever consider developing."

Researchers working on their own differentiation projects have to be willing to perform the same experiments repeatedly. "The most important thing is to have very good control of the culturing conditions," Baetge says. "Don't keep the ES [embryonic stem]cells around forever. Make a large bank of cells in vials. Use a vial to do the experiment and when it's done, don't keep passaging the cells. Thaw a new vial and begin again."

## Taking account of proteins


Five years ago in our anniversary issue, proteomics experts told us that they dreamed of the day when entire proteomes could be displayed and meaningful cross-platform and cross-laboratory comparisons could be made. The field is now closer to realizing that goal, according to Christine Vogel, who in 2007, with her then post-doctoral advisor, Edward Marcotte, and colleagues, reported a method for large-scale quantification of proteins. Previously, quantifying proteins could only be done on a relative scale, and relied on laborious, impractical and sometimes expensive labeling methods. Marcotte and his colleagues at the University of Texas, Austin, moved the field toward a label-free approach with a technique they call APEX (absolute protein expression profiling)[29].

Edward Marcotte says, "It's clear that ability to monitor proteomes allows you to look directly at all sorts of things."

APEX works by taking a count of peptides (so-called spectral counting) and then adjusting that value by the likelihood that a particular peptide will be present, or by what Vogel calls its "flyability" (called by others "frequent flyers"), which is a function of various things, such as ionization efficiency and solvent conditions. Using this method in conjunction with expression profiling, they

Christine Vogel says, "APEX has helped establish quantitative protein concentration as a novel data type."

showed for several thousand yeast and *Escherichia coli* proteins that a protein's abundance correlates most of the time with the amount of message present—70% of the time with yeast, slightly less with *E. coli*. This work dovetailed nicely with a paper from Ruedi Aebersold's group in the same issue, in which a computational approach was taken to identify some 16,000 proteotypic peptides (those with the greatest flyability, most likely to be detected by mass spectrometry) for over 4,000 yeast proteins[30].

Since then, APEX has become a workhorse to address a range of research questions in the Marcotte laboratory (and will soon in the Vogel laboratory, which is in the processing of being set up at New York University). "It was fun for us. It opened up a lot of studies," says Marcotte. "It's clear that ability to monitor proteomes allows you to look directly at all sorts of things, organization of proteins into complexes, the function of proteins. We still don't know the function of a large number of proteins. It will allow us to nail down function. Think of metaproteomics, [APEX can reveal] which gene products survive in different niches."

More recent work from the Marcotte laboratory and elsewhere has shown that the link between mRNA levels and protein levels is not as simple as it appeared in yeast and *E. coli*. Applying APEX to a human tumor cell line, for example, Marcotte found that you can explain only 30% of the variation in protein concentrations with the knowledge of mRNA concentrations in the same cell, which means that over two-thirds is a function of post-transcriptional regulation of some kind[31]. Not only that, in the hands of a group of Swiss researchers[32] as well as Marcotte[33], APEX has been used to show that protein abundances across species from mice to man show greater conservation than RNA levels. Marcotte likes to think about it in this way: "RNA abundances are free to diverge, while post-translational regulation brings protein levels back in line."

Since the paper came out, APEX has been accepted as a fast and easy method for getting absolute quantity. Label-free techniques open the door for experiments where labeling would be impossible or when it might change an organism's metabolism or a protein's properties. "Along with work by others, [the *Nature Biotechnology* paper] helped establish the prediction of flyability, which, although not perfect, is usable for everyday life, and has helped establish quantitative protein concentration as a novel data type," says Vogel.

APEX has been taken up by two groups of researchers—method developers ("the MS [mass spectrometry] crowd") and biologists—according to Vogel, who puts herself in the latter group. "It points the researchers to the usefulness of absolute concentrations and our ability to now measure them at large scale. Absolute concentrations are needed to start thinking about 'rates' of protein production and degradation. Several papers have appeared which start thinking about translation rates, and such models of rates."

And as for the dreams of five years ago, several have come true. Researchers can now routinely identify proteins at large scale and quantify them using ICAT (isotope-coded affinity tags), SILAC (stable isotope labeling with amino acids in cell culture) or other labeling methods, as well as label-free strategies. Moreover, the Aebersold laboratory has created a proteome-wide map of proteotypic peptides for yeast and (to a large extent) for human[34].


1. Takahashi, K. & Yamanaka, S. *Cell* **126**, 663–676 (2006).
2. Nakagawa, M. *et al. Nat. Biotechnol.* **26**, 101–106 (2008).
3. Miura, K. *et al. Nat. Biotechnol.* **27**, 743–745 (2009).
4. Nakagawa, M. *et al. Proc. Natl. Acad. Sci. USA* **107**, 14152–14157 (2010).
5. Shojaei, F. *et al. Nat. Biotechnol.* **25**, 911–920 (2007).
6. Shojaei, F. *et al. Nature* **450**, 825–831 (2007).
7. Shojaei, F. *et al. Proc. Natl. Acad. Sci. USA* **106**, 6742–6747 (2009).
8. Ball, M.P. *et al. Nat. Biotechnol.* **27**, 361–368 (2009).
9. Deng, J. *et al. Nat. Biotechnol.* **27**, 353–360 (2009).
10. Choi, H.S. *et al. Nat. Biotechnol.* **25**, 1165–1170 (2007).
11. Choi, H.S. & Frangioni, J.V. *Mol. Imaging* **9**, 291–310 (2010).
12. Qian, X. *et al. Nat. Biotechnol.* **26**, 83–90 (2008).
13. Mohs, A.M. *et al. Anal. Chem.* **82**, 9058–9065 (2010).
14. Fabian, M.A. *et al. Nat. Biotechnol.* **23**, 329–336 (2005).
15. Karaman, M.W. *et al. Nat. Biotechnol.* **26**, 127–132 (2008).
16. Manning, G. *et al. Science* **298**, 1912–1934 (2002).
17. Goldstein, D.M., Gray, N.S. & Zarrinkar, P.P. *Nat. Rev. Drug Discov.* **7**, 391–397 (2008).
18. Shi, L. *et al. Nat. Biotechnol.* **24**, 1151–1161 (2006).
19. Shi, L. *et al. Nat. Biotechnol.* **28**, 827–838 (2010).
20. Duarte, N.C. *et al. Proc. Natl. Acad. Sci. USA* **104**, 1777–1782 (2007).
21. Shlomi, T., Cabilil, M.N., Herrgård, M.J., Palsson, B.Ø. & Ruppin, E. *Nat. Biotechnol.* **26**, 1003–1010 (2008).
22. Zur, H., Ruppin, E. & Shlomi, T. *Bioinformatics* **26**, 3140–3142 (2010).
23. Zheng, G., Patolsky, F. & Cui, Y. Wang, W. U., & Lieber, C. M. *Nat. Biotechnol.* **23**, 1294–1301 (2005).
24. Cui, Y., Wei, Q., Park, H. & Lieber, C.M. *Science* **293**, 1289–1292 (2001).
25. Tian, B. *et al. Science* **329**, 830–834 (2010).
26. D'Amour, K.A. *et al. Nat. Biotechnol.* **24**, 1392–1401 (2006).
27. D'Amour, K.A. *et al. Nat. Biotechnol.* **23**, 1534–1541 (2005).
28. Kroon, E. *et al. Nat. Biotechnol.* **26**, 443–452 (2008).
29. Lu, P., Vogel, C., Wang, R., Yao, X. & Marcotte, E. *Nat. Biotechnol.* **25**, 117–124 (2007).
30. Mallick, P. *et al. Nat. Biotechnol.* **25**, 125–131 (2007).
31. Vogel, C. *et al. Mol. Syst. Biol.* **6**, 400 (2010).
32. Schrimpf, S.P. *et al. PLoS Biol.* **7**, e48 (2009).
33. Laurent, J.M. *et al. Proteomics* **10**, 4209–4212 (2010).
34. Picotti, P. *et al. Cell* **138**, 795–806 (2009).