

digital and data officer, Milind Kamkolkar. He adds that, in many cases, computational experts operate in relative isolation from the clinicians, chemists and biologists with the deepest grasp of the disease being studied. “The informatics team is often the last team to understand the scientific hypotheses or why they’re looking at this particular data,” says Kamkolkar.

This new generation of machine-learning-assisted drug discovery startup now aims to overcome these challenges through tight integration of computation and benchwork, with close collaboration between both sets of experts. On-site data generation is a critical component of this process. “We built our own biology lab, where we do protein production, structural genomics, biophysical screening and pharmacology,” says Andrew Hopkins, founder and CEO of Exscientia. “For many projects, we are the first people to screen against a given target.”

The data being generated vary from company to company. LabGenius is focused on protein engineering efforts, with experimental processes that can identify new protein therapeutics and optimize the biochemical and pharmacological properties of existing proteins. Recursion’s strategy emphasizes cell and tissue imaging, and the company claims to have assembled the world’s largest collection of such data. Other companies, like Insitro and Celsius Therapeutics, are relying on single-cell transcriptome analysis to understand how different cell types behave in disease and health, or in response to specific perturbations. “We have a high-throughput biology lab that is essentially a factory for data, using human-derived induced pluripotent stem cells that span a spectrum of human genetic diversity,” says Insitro’s Koller, noting that they make extensive use of CRISPR technology to test the effects of known disease-related mutations and also work with relevant donor-derived specimens.

These results are then fed into analytical pipelines. Some of the goals here are relatively conventional in the computational biology world — for example, classifying distinct cell types or states on the basis of gene expression profiles. But these algorithms can also be used to assemble more extensive descriptions of cellular states that incorporate multiple data modalities such as imaging, genomics, transcriptomics and proteomics into a single, detailed representation. Cellarity calls such representations ‘maps’, which can in turn be used to identify signposts of healthy or abnormal function. “We can’t possibly understand every little detail of what’s

going on everywhere,” says chief technology officer Chad Nusbaum. “We’re looking for biomarkers of the behavior of that cell in a healthful and an unhealthful state, and to predict what’s going to encourage the cell to go to the healthful state.”

Although potentially informative, such models are just educated guesses until experimentally validated. But rather than testing their models with massive high-throughput screens that may ultimately prove fruitless, these startups again leverage machine learning to plan next steps.

“Active learning is about asking which compounds I should make and test that would be most informative to move my project forward,” says Hopkins. “It’s all about data acquisition.” Rather than thousands or more of compounds, these algorithms might propose a far smaller number of experiments, including drug candidates predicted to fail as well as those expected to succeed. This makes active learning a potent hypothesis-generation and testing engine that can help drug developers focus their efforts. “Eliminating the ‘bad apples’ early is where I see the main advantage of active learning,” says Schneider. The proposed experimental plan then goes back to the experimentalists at the bench, kicking off a new cycle of testing and analysis.

This iterative development is an essential component of the process. “AI-based drug design is not going to just give you a drug straight off — that’s still science fiction,” says Hopkins. At Celsius, for example, researchers carefully analyze the distinct cell types and gene expression profiles identified in initial single-cell transcriptomic experiments and then test those models by using tools like CRISPR to introduce targeted genomic manipulations that might confirm or invalidate potential drug targets in disease-affected cells. “It helps generate these hypotheses faster and then test them — it’s a lot faster to push electrons around than cells,” says chief data officer Gregory Ryslik. And as the algorithms and workflows improve, the process can achieve greater efficiency; Hopkins notes since Exscientia’s launch in 2012, the company has streamlined the number of drug leads typically needed to obtain a candidate suitable for clinical development from around 2,500 to 400 or fewer.

These gains can be notably accelerated if automation is brought into play. In this scenario, the actual design of new experiments still falls to human scientists, but once the workflow is formalized and validated, it can easily be automated. At Insitro, Koller notes that they can now perform experiments within several weeks that would take years for most academic

First drug approved for neurofibromas is a MEK inhibitor

AstraZeneca found a winning indication for Koselugo (selumetinib), the first drug to be approved by the US Food and Drug Administration (FDA) for neurofibromatosis type 1 (NF1), a genetic disorder that causes benign tumor growth in nerves. Koselugo is approved for pediatric patients 2 years of age and older with symptomatic, inoperable **plexiform neurofibromas** — nerve sheaths tumors — which can grow anywhere in the body, including the spine and organs. Koselugo is a small molecule that targets the kinase MEK, which phosphorylates mitogen-activated protein kinase (MAPK). The drug’s approval rests on the back of a phase 2 trial conducted at the US National Cancer Institute. **Trial results** showed a partial response in all 50 patients and an overall response rate of 66%, most responses lasting 12 months or longer.

NF1 is caused by germline mutations in the **NF1 tumor suppressor gene** and affects 1 in 3,000 people. It is characterized by elevated RAS–MAPK signaling. Koselugo is an oral selective inhibitor of MEKs 1 and 2 and the fourth MEK inhibitor approved overall (the other indications are melanoma and non-small-cell lung cancer). The FDA previously granted AstraZeneca an Orphan Drug designation for Koselugo in pediatric NF1, as well as Priority Review and Breakthrough Therapy designations. The drug has been tested in clinical trials for at least 13 tumor types.

Mutations in the NF1 gene are known drivers in multiple cancers. AstraZeneca still has Koselugo in phase 1 testing in combination with Imfinzi (durvalumab) for solid tumors and in phase 2 testing in combination with Tagrisso (osimertinib) for advanced non-small-cell lung cancer expressing the epidermal growth factor receptor. The company tested Koselugo with docetaxel as a second-line treatment for KRAS-positive NSCLC, but the combination failed to meet its primary endpoint of progression-free survival in a phase 3 trial.

Published online: 11 May 2020
<https://doi.org/10.1038/s41587-020-0530-3>