From Cerebral Cavernous Malformation to cancer research:
A one-size-fits-all path to precision medicines

By Adam Chernick, PhD
“Being a cancer survivor has helped me appreciate life more and be grateful for those everyday things that are easy to take for granted,” says Chris. “Since my diagnosis, I’ve started golfing, which has become my new passion. When I’m not golfing, I’m at the beach. I’m like a kid: I’m not indoors until the sun comes down. I don’t want to miss a thing! I will always have cancer, but I am living life to its fullest.”

Introduction and background
Cancer, like many common diseases, is a complex, multifactorial condition. It affects approximately 18 million people globally and with 1 in 6 deaths occurring due to cancer, it is the second leading cause of death worldwide. Despite this, cancer therapies have the lowest success rate in clinical trials compared with all the other major diseases. This is partly due to the lack of understanding regarding many aspects of this disease, including its underlying pathological mechanisms and treatment targets. These knowledge gaps hinder the development of better treatment options for patients, especially precision oncology medicine. Key unanswered questions include how can the best treatment option be identified for individual patients? Can artificial intelligence (AI) be used to detect cancer earlier? And why do certain faulty genes only cause cancer in certain parts of the human body?

To help answer these and other questions, Recursion, the digital biology company, is combining biological imaging with machine learning (a form of AI) and CRISPR, to decode the complex biology of cancer and other multifactorial diseases, such as neurofibromatosis type 2 and GM2 gangliosidosis. Using CRISPR, every gene in the human genome is systematically knocked out, one by one, and the cellular phenotype of each knockout is captured using a proprietary staining protocol consisting of six subcellular dyes imaged in six channels using high resolution fluorescent microscopy. Machine learning tools are trained to identify the most salient features of images and aggregate cellular response induced by CRISPR perturbations and quantify these changes in an unbiased manner. This data is used to create an entirely relatable “phenomic” map of the human genome and generate multiple disease models, which can be used to not only elucidate pathological processes but also to test potential treatments of disease. And as many of the iterative processes involved are automated, Recursion is effectively industrializing drug discovery.

In collaboration with Integrated DNA Technologies (IDT), Recursion has applied hundreds of thousands of CRISPR guides in under a year.

Cerebral Cavernous Malformation
Recursion’s novel approach to drug discovery was inspired by the concepts and methods initially developed for a doctoral project on cerebral cavernous malformations (CCM), a hereditary hemorrhagic stroke syndrome. Cerebral cavernous malformations are abnormally formed capillaries in the brain that are enlarged and irregular in structure. These blood vessels have abnormally thin walls and are prone to leakage. They also lack supportive tissues such as elastic fibers, which reduces the elasticity of these capillaries. This means that the capillaries may not return to their original size and shape after filling up with blood, resulting in the formation of caverns. While 25 percent of people with CCMs never experience any related health problems, others may experience serious neurological symptoms associated with cerebral hemorrhage, that is, bleeding in the brain. These include headaches, seizures, weakness or numbness on one side of the body, paralysis, difficulty speaking, difficulty understanding others, and hearing or vision loss. Recurrent bleeding can result in progressively worsening symptoms, and severe hemorrhages can be fatal.

Although the majority of CCM cases are sporadic and idiopathic, at least 20 percent of all cases are familial. Familial CCM is inherited in an autosomal dominant pattern and is caused by monogenic loss-of-function mutations in any one of the three genes identified so far, KRIT1 (also known as CCM1), CCM2, and PDCD10 (also known as CCM3). Similar to cancer, a Knudsonian two-hit loss-of-heterozygosity hypothesis has been proposed for the etiology of CCM, requiring that both alleles of a gene be ‘hit’ before the disease fully manifests. These hits may occur as two separate mutation events, where one hit may have been inherited and the other hit incurred during life. Together, there may be as many as 360,000 patients with symptomatic familial and sporadic CCM in the US and EU5.

A bold new path to repurposing drugs
In order to identify known drugs that could be repurposed to treat CCM, an early predecessor to Recursion’s current unbiased screening platform based on cellular models of CCM due to CCM2 loss-of-function was developed, with additional validation in animal models. This discovery strategy involved four steps. One, drugs were screened using a cell culture model of CCM created by the transfection of primary human adult cells with small interfering RNAs (siRNA) to knock out either the CCM2 gene or a scrambled control gene. Any effects produced by the drugs were identified by microscopic imaging of the fixed cells, which had been stained using fluorescent probes. Cellular structures, including the nucleus, actin stress fibers and VE-cadherin cell–cell junctions, were measured and any drugs that rescued the disease phenotype in these cells were selected for further validation. Two, drugs were screened a second time, using a method orthogonal to the first. This trans-cellular resistance screen involved the real-time measurement of electrical resistance encountered when passing an electrical current through a monolayer of cells. The cells used were again those containing siRNA knockouts of either CCM2 or a scrambled control. CCM2-deficient cells displayed a functional defect in their monolayer stability, which was detected using trans-cellular resistance. Drugs that demonstrated rescue of the functional defect were then prioritized for further investigation, using animal models. Three, CCM2 knockout mice were injected with intradermal wheals of the drugs selected in the first two screens. The leakage of capillaries around the injection sites 30 minutes after injection were measured. Any drugs that reduced the peri-injection microvascular leakiness in this acute animal model of CCM were then assessed in a chronic animal model of CCM. Four, the therapeutic potential of the drugs for CCM was measured using magnetic resonance imaging (MRI) of brains from CCM2 knockout mice that had been treated with the drugs for 5 months since their fifth day after birth. The number and size of CCM lesions were measured and any drugs that reduced the number and size of the lesions were considered as potential candidates for progression into further preclinical testing.

Man vs machine
Other than identifying two compounds that could potentially treat CCM, out of 2,100 initial candidates, this academic study demonstrated the
proof of principle behind the unbiased, phenomic screening approach. A critical finding of the study was also revealed when comparison was made between results obtained using human qualitative analysis and those obtained using software that incorporated machine learning. In the first screen, the various effects of the drugs on the cell culture model of CCM were measured using CellProfiler and CellProfiler Analyst software, for automated high-throughput screening of the microscope images. CellProfiler is an imaging analysis tool, developed by the Broad Institute of MIT and Harvard, which imported images, identified the borders of each cell, and created a database of a multitude of mathematical descriptors of every cell in every image collected. CellProfiler Analyst is a machine learning tool, which was employed to develop rules that could be used to distinguish whether each cell in an image was more likely to have been treated by siRNA targeting CCM2 or a scrambled control. The software accurately characterized the images and thus, identified several drugs as potential therapeutic options for CCM because they rescued the disease phenotype.

Simultaneously, the same microscope images were analyzed and qualitatively scored by two human reviewers, who identified 38 drugs that rescued the cellular disease phenotype and were therefore, potential treatments for CCM. Interestingly, there was no overlap between the 38 drugs selected by the humans and the top 38 drugs ranked by the software as potential CCM therapies. Of the 38 identified by the humans, only one drug was validated by the second screen, while seven drugs identified by the automated machine-learning analysis demonstrated functional rescue of the cellular disease phenotype. As the remaining drug identified by human analysis, simvastatin, had already failed to demonstrate a reduction of CCM disease burden in CCM2 knockout mice, further investigation of this drug was not pursued. Two of the seven identified by the automated analysis, cholecalciferol (also known as vitamin D3) and tempol, continued to be validated in the last two screening assays (steps 3 and 4). Not only did the automated characterization and AI-guided analysis provide more accurate results than that produced by humans, it also provided new

<table>
<thead>
<tr>
<th>Disease Model</th>
<th>Early Discovery</th>
<th>Discovery</th>
<th>Preclinical</th>
<th>Phase 1</th>
<th>Phase 2</th>
<th>Phase 3</th>
<th>Commercial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neurofibromatosis Type 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cerebral Cavernous Malformation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GM2 Gangliosidosis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solid Tumors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dilated Cardiomyopathy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clostridium difficile Colitis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neuroinflammation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Batten Disease</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Charcot-Marie-Tooth Disease</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pulmonary Arterial Hypertension</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Senescence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10+ Early Discovery Programs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 1: The Recursion pipeline of drug discovery and development programmes.12
avenues of investigation by identifying drugs that work through biological pathways previously not known to associate with the disease. The success of the automated machine-learning tools was attributed to its unbiased approach and ability to detect more subtle phenotypic changes than humans, who are hampered by preconceived bias and inattentional blindness.7

**Industrializing drug discovery**

The approach, strategy and methods conceived and developed in this academic study has formed the foundation of the drug discovery and development platform at Recursion. This study was completed in 2013 and the company was founded the same year. Since then, Recursion has been developing a treatment for CCM referred to as REC-994. REC-994 received clearance as an investigational new drug (IND) from the US Food and Drug Administration (FDA) in 2018, enabling it to enter and complete Phase 1 clinical studies for the treatment of CCM.11 This is currently just one of four clinical-stage therapeutics that Recursion have discovered and developed (see Figure 1).12 Another of those clinical-stage compounds is REC-4881, a mitogen-activated extracellular signal-regulated kinase (MEK)-inhibitor. REC-4881 is being developed to reduce tumor size in familial adenomatous polyposis, or FAP, a hereditary tumor syndrome. Supported by prior positive pharmacokinetic, pharmacodynamic and Phase 1 clinical data, REC-4881 is in preparation for a Phase 2 clinical trial to evaluate its efficacy and safety for the treatment of FAP. In hereditary tumor syndromes, certain patterns of cancer may be observed among families, with several first-degree relatives developing the same type of cancer, or developing cancer at an early age, or having two or more types of cancer develop in the same person.13 Alongside the discovery and development of new drug candidates, Recursion have continued to refine and improve their drug discovery platform. This includes continuing to train the machine learning tools and integrating novel genetic tools like CRISPR gene editing. The CRISPR technology has proved to be much more precise and specific, with fewer off-target effects than prior work with siRNA. The much cleaner signal not only means that their platforms now function more accurately but also more robustly. To date, the company has accrued nearly 7 petabytes of data, a substantial proportion of which – about 1 PB – they have made publicly available. This includes the release of RxRx2, which has made open access their entire dataset on the SARS-CoV-2 virus.14,15 This exemplifies the advantages of the platform to decode complex biology, enabling the rapid elucidation of pathological states through highly accurate and comprehensive phenomic profiling of multifactorial diseases. The platform can be deployed to analyze a multitude of diseases simultaneously and is not limited by the constraints associated with more traditional approaches to drug discovery and development, where usually only a handful of diseases or treatment targets are investigated. Moreover, the unbiased target-agnostic approach means that the platform will just as easily identify and validate novel treatment targets as consolidate evidence to support existing targets. Its broad application, high throughput, speed, and computational power has led to the initiation of more than 30 early discovery programmes and the achievement of four lead compounds entering Phase 1 or 2 clinical trials, that is double the number of clinical-stage assets since last year (see Figure 1).

**Mapping the human phenome**

Another big project that is underway is the phenomics map of the human genome. This involves knocking out every human gene, one by one, and characterizing the phenomic profile of each knockout in diverse human cell types using a cellular phenomics assay. The genes are precisely knocked out using CRISPR-Cas9. To keep pace with the large-scale needs of this project, Recursion have collaborated with IDT to design and obtain over 100,000 CRISPR guides. This has enabled the phenomics project to progress rapidly toward mapping the entire human genome. Although Recursion does design their own CRISPR guides, the availability of guides and other CRISPR technologies from IDT has vastly improved the high-throughput capability of their platforms, eclipsing what was achievable even last year. On its completion, the phenomics map will facilitate the interrogation of disease biology to identify treatment targets and biomarkers for disease profiling and treatment response at breakthrough pace. For example, in a first project, the Recursion oncology team identified a phenomics-inferred small molecule and demonstrated proof of concept efficacy in a mouse model of cancer in 3 months. Known drugs can be screened to repurpose them for more clinical indications, improving the utility of marketed drugs as well as orphan drugs. This can and should of course include precision medicine, and is in the case of REC-4881, advancing towards precision oncology. The platform has the potential to identify the best treatment options available to treat that particular type of tumor in these individuals based on their genetic profile. Thus, Recursion are harnessing powerful analytical tools, such as machine learning and CRISPR, in a bold modern way, to industrialize drug discovery and development to treat patients with a broad spectrum of disease, ranging from relatively rare monogenic disorders, such as CCM, to common multifactorial diseases, such as cancer.14

---

**References**

* Quote is a composite based on quotes from real patients, Chris P. and Anthony B., shared at https://www.cancercenter.com/patient-stories/.

---

**Disclaimer**

© 2021 Integrated DNA Technologies. All rights reserved. Trademarks contained herein are the property of Integrated DNA Technologies, Inc. or their respective owners. For specific trademark and licensing information, see www.idtdna.com/trademarks. Any IDT products referenced in this article are for research use only. IDT does not intend these products to be used in clinical applications and does not warrant their fitness or suitability for any clinical diagnostic use. Any purchaser of these products is solely responsible for all decisions regarding the use of these products and any associated regulatory or legal obligations.