Biotechnology

BioInsights Virtual Series: Takeaways from Our AI/ML Panel & 13 Fireside Chats

CONCLUSION

Last Friday, we hosted a series of fireside chats and a panel discussion for our BioInsights Virtual Series: AI & Machine Learning Powers the Future of Drug Discovery. During this event, we hosted 13 different fireside chats with executives from several key players in the field, including: **Deep Genomics** (Private), **Excision BioTherapeutics** (Private), **Exscientia PLC** (EXAI, not covered), **Generate Biomedicines** (Private), **Immuneering** (IMRX, Malekar), **Insitro** (Private), **Metagenomi** (Private), **Odyssey Therapeutics** (Private), **Recursion Pharmaceuticals** (RXRX, not covered), **Schrödinger** (SDGR, Kim), **Shape Therapeutics** (Private), **Totus Medicines** (Private), **XtalPi** (Private). Our Panel discussion featured executives from **Celsius Therapeutics** (Private), **Omega Therapeutics** (OMGA, Tenthoff), **Scorpion Therapeutics** (Private), **Systems Oncology** (Private), and **Valo Health** (Private).

- **Harnessing AI and Machine Learning in Designing Better Therapies.** The panelists agreed that we are still in the early days of AI/ML-empowered drug discovery and development. While each of the companies on the panel have different approaches and are applying AI/ML in different ways, all are employing AI/ML to develop new and better drugs for patients faster. Some companies view AI/ML as a front-end drug discovery tool, while others have integrated AI/ML throughout the development process to enhance clinical trial design and development. Unprecedented computational power is enabling even small companies to analyze enormous data sets unlike ever before. Integration of data and iteration are keys to successfully exploiting AI/ML. As Dr. David Berry, CEO of Valo stated, "We are on the precipice of the digital transformation of pharmaceutical R&D." Tarriq Kassum, CEO of Celsius said, "AI/ML enables us to solve tougher problems that we haven't been able to solve before." Clearly AI/ML will play a key role in dramatically accelerating and enabling the discovery and development of future medicines.

Yasmeen Rahimi, Ph.D.
Sr. Research Analyst, Piper Sandler & Co.
646 819-3586, yasmeen.rahimi@psc.com

Edward A. Tenthoff
Sr Research Analyst, Piper Sandler & Co.
212 284-9403, edward.tenthoff@psc.com

Do Kim
Sr. Research Analyst, Piper Sandler & Co.
212 284-9488, do.kim@psc.com

Christopher J. Raymond
Sr. Research Analyst, Piper Sandler & Co.
312 267-5086, christopher.raymond@psc.com

Swapnil Malekar, Ph.D.
Research Analyst, Piper Sandler & Co.
646 531-6851, swapnil.malekar@psc.com

Related Companies:	Share Price:
GILD	60.67
IMRX	7.81
MRNA	172.54
OMGA	7.19
SDGR	35.90

INDUSTRY RISKS

Companies may fail to achieve development, regulatory, or commercial objectives.

Panel: Harnessing AI and Machine Learning in Designing Better Therapies

We hosted a CEO panel with Mahesh Karande of Omega Therapeutics (OMGA, Tenthoff), Dr. David Berry, M.D., Ph.D. of Valo Health, Dr. Tariq Kassum, M.D. of Celsius Therapeutics, Dr. Axel Hoos, M.D., Ph.D. of Scorpion Therapeutics and Spyro Mousses, Ph.D. of Systems Oncology. While each of the companies have different approaches and are applying AI/ML in different ways, all are employing AI/ML to develop better drugs for patients faster. Here is what we learned:

Companies are using AI/ML to analyze large data sets. Unprecedented computational power is enabling even small companies to analyze enormous data sets unlike ever before. As David said, "We are at a new scale of data compute power allowing us to do things at a different scale and level of integration." Omega used AI/ML and proprietary computational algorithms to identify and map >15,000 insulated genomic domains (IGDs) across the human genome, and even identified specific binding sites called *EpiZips* within these IGDs.

Celsius is a precision immunology company focused on autoimmune diseases and oncology. The company's *SCOPE* (Single Cell Observations for Precision Effect) platform performs single cell RNA sequencing on >1,000 curated human tissue samples generating large quantities of high dimensional data. Celsius employs AI/ML to clean-up raw data sets and uses custom algorithms based on topic modeling inspired by natural language processing to identify gene expression profiles that correlate with diseases. Scorpion is a precision oncology company that combines AI/ML with molecular biology and medicinal chemistry to design better cancer drugs.

Integration of data and iteration are keys to successfully exploiting AI/ML. Systems Oncology employs Hypergraphs to link different data types using set theory. Systems Oncology is mining data to target 3-4 pathways simultaneously to screen **100 quadrillion** (1x10^20) possible combinations to identify synthetic lethality in different cancer types!

Valo is regarded as the first fully-integrated digital pharmaceutical company. OPAL is an integrated AI/ML architecture that spans the entire drug discovery and development process. OPAL contains high density, longitudinal, clinical and multi-omic data on ~8 million patients over an average of 15 years. Not only can OPAL elucidate causal biology, but also identify small molecules and predict likely safety, drug-drug interactions, and efficacy in the clinic. Additionally, OPAL can identify biomarkers for patient selection to design dramatically smaller clinical studies that will take less time and cost with higher likelihood of success. Importantly, OPAL is iterative, such that the platform gets better and faster with each cycle and more data.

The ultimate goal is to discover/develop new and better drugs. Some companies view AI/ML as a tool to apply to drug discovery to identify better drugs to take into the clinic. AI/ML can be used not only to identify more selective drugs against existing targets, but also to identify drugs against undruggable targets.

Mahesh pointed out that AI/ML can also be applied to create entirely new medicines and novel therapeutic modalities. Omega is developing *Omega Epigenetic Controllers (OECs)* to precisely control the level and duration of expression of genes to treat a broad range of diseases. Omega uses AI/ML to design mRNA strands that encode for a fusion protein comprised of a DNA-binding protein targeting a specific EpiZip within a selected IGD, and an effector protein that modulates expression of a target gene. OTX-2002 is designed to epigenetically modulate c-MYC for liver cancer with data at AACR and an IND filing planned in 1H22.

Systems Oncology is developing a novel therapeutic modality called RNA *SeekRs* that contain several siRNAs to simultaneously silence multiple disease targets conjugated to two aptamers to specifically deliver to targeted cancer cells. The company will report preclinical data at AACR on its lead colon cancer program, which is comprised of aptamers targeting EpCam and Her3 to deliver siRNAs against multiple ubiquitin family members. Systems Oncology's 2nd program will also to be highlighted at AACR, targeting PD-1 and CTLA-4 to deliver multiple siRNAs, including one against a transcription factor involved in T-cell exhaustion.

Other companies have integrated AI/ML throughout the drug discovery/development process to not only identify new drugs, but also use the data to enhance clinical development. As Tariq put it, "AI/ML enables us to solve tougher problems that we haven't been able to solve before." In developing lead program CEL383, an anti-TREM1 antibody for inflammatory bowel disease (IBD), Celsius used AI/ML to identify the cell type that correlated to inflammation and TNF-resistance, TREM1 target, and a patient subset who are expected to respond. Celsius will file an IND for CEL383 and initiate the company's first Phase I study in early 2023.

Valo is developing drugs to treat cardio/metabolic/renal and neurodegenerative diseases, plus oncology. Because CV pipelines are thin and traditionally the area of Pharma due to large expensive clinical trials, Valo believes OPAL's ability to cut time and cost by enabling smaller trials could be especially productive. Valo is conducting a Phase II study of OPL-0301, an S1P1R agonist, for left ventricular dysfunction following myocardial infarction (MI) with future potential applicability in heart failure and acute kidney injury. OPAL was able to predict that OPL-0301 would not interact with standard-of-care beta blockers. Further, OPAL identified a biomarker that enables patient selection to decrease the Phase II study size by 80% and a 3-month surrogate efficacy endpoint that will dramatically reduce time to PoC data. If successful, Valo believes OPAL could cut 7 years and 80% of the cost of developing OPL-0301 vs. traditional CV drugs! Valo is also preparing to initiate a Phase II study of oral OPL-401 for diabetic retinopathy. Using OPAL, Valo developed a proprietary algorithm based on structural changes in the back of the eye and visual acuity able to predict those diabetic retinopathy patients who will progress, so that OPL-0401 studies can be smaller and achieve PoC faster.

Some noteworthy thoughts and conclusions. The panelists agreed that we are still in the early days of AI/ML-empowered drug discovery and development. As Spyro put it, "We are starting to approach where machines can do as well as humans. The future will be going beyond what humans can do to discover new biology." David stated, "We are on the precipice of the digital transformation of pharmaceutical R&D.…Digital transformation of other industries has led to step function changes." David added that it will "take a commitment to data, to data integration, to transforming the use of data, investing in analytical tools to extract out conclusions and then substantially apply those conclusions." Clearly AI/ML will play a key role in dramatically accelerating and enabling the discovery and development of future medicines.

Deep Genomics

Abundance of RNA-biology data has allowed for development of a robust AI/ML platform. While rationalizing their approach to initially utilize RNA therapies, mgmt noted that previous RNA therapies have revolutionized the filed to a certain extent and cited the examples of Spinraza and COVID vaccines. It was noted that one can program the oligonucleotide sequence (by changing the letters) to increase or decrease the target gene expression. Furthermore, the RNA biology, DNA sequences, and gene sequences were noted to be present in a digital format that can be easily utilized as snippets to alter the target expression and achieve desired biological effect. Mgmt also noted that there is an abundance of data pertaining to RNA biology that can be utilized to strengthen the company's platform and increase its predictive capabilities. Additionally, Deep Genomics' AI/ML platform has the ability to be modality agnostic to a certain extent and can be utilized to develop ASO, siRNA, gene therapy products, and DNA editing tools. It was noted that RNA biology is the foundation of their platform that allows them to measure relevant information that are good predictors of the output. Additionally, single cell sequencing was noted to be a very useful tool where data can be measured rapidly. Finally, the experience of the team was emphasized to generate quality data and noted that the CEO has been in this field for 20 years along with significant experience of other members in the organization that contributes to the team's success.

What makes Deep Genomic's platform different from others? It was noted that while generally thinking about AI companies 5 ingredients should be kept in mind including: **(1) Framework:** A company needs to have framework of actionable items and leads to predictive results and that Deep Genomics has such a system in place where extensive amount of RNA biology related info is input to achieve quantifiable output; **(2) Prediction:** Deep Genomics uses data to make predictions with AI/ML and mentioned that when looking at a new company,

investors should be looking at the predictive power of the platform and how useful it is to discover novel targets/molecules; **(3) Data:** It was noted that people always have lots of data on hand and do not know what to do with it (for eg: electronic health record data is all over the place and needs to be fed in a proper framework to drive meaningful conclusions) and that Deep Genomics is sitting on 100 petabytes of data that can be input into their well-designed framework; **(4) Positive feedback:** Mgmt noted that companies should make predictions using their platform and use the data output to feedback into the framework to improve the ML predictors. Additionally, it was noted that if it takes 4 yrs for feedback, it indicates a slow progressing module while Deep Genomics' feedback happens within 1 month; and **(5) Unlimited Power:** Mgmt commented that the AI platform should "surprise" inventors by opening new doors and unlocking new things, and explained that they have built predictive tools that can analyze mutations to discover novel targets for therapies. Same AI system can be used to develop mouse models (can introduce human mutations in mouse) and utilization of large databases like the UK BioBank to discover novel targets.

What are the barriers in RNA biology discovery and how does AI/machine learning overcome these limitations? It was noted that most of the people are focused on speed and cost of drug discovery which leads to hurdles and one needs to think about quality of the programs to ensure a high POS. For example, one of Deep Genomics' programs is drugging "undruggable targets" where others have tried and failed in the past and their AI platform looked at 640k novel compounds (oligo) that led to prioritization and design of 362 olignucleotides across 7 different MOAs to express the gene of interest. Furthermore, when tested in labs, 96 compounds were validated reversing the target gene expression by 50%. It was also noted that to achieve a similar result, traditional approaches would have to screen 100x more compounds making the process longer and expensive. Mgmt also noted that the merit of AI/ML platform should remain relevant when a new technology is introduced (and the existing capabilities does not become obsolete). To that extent, Deep Genomics has the capability to integrate new tech into its platform that allows for faster screening. They are also equipped with capabilities that can screen for combination effects of 2 oligos where twice the processing power is required. Mgmt also pointed out that same capability can be used to assess the knockout vs increase in expression of a gene as the regulatory elements in a gene are uniform and can be targeted specifically to achieve a desired effect. Mgmt also noted that there are different types of toxicities associated with oligos such as non-specific Watson-Crick binding or lack of base pair matching.

Excision BioTherapeutics

What is the novelty behind Excision's AI/ML approach to developing CRISPR based therapies to cure viral diseases? Excision uses a range of computational methods during each step of the pipeline and utilizes machine learning to robustly build, test, and refine the technology. They note that the board companies in this space are well versed in computational methods of complex data sets, but at Excision they are all in on targeting virus and look at conservation, entropy, targeting strategies and ensuring specificity to match patient profiles and effectively target those sequences. It was noted that from the first step of looking at viral databases to help in choosing a guide RNA, this step ensures highly conserved and very specific target sites for their viral targeting. As stated, when one is targeting gene correction, one tends to cut very close to the disease-causing mutation and therefore the choice of guides is limited. Particularly when targeting the human genome for a knock-out, the target coding sequence is generally similar to other spots in the genome. With the computational methods that Excision is using to determine a larger range of sites within the virus that can be targeted that are very different from the human genome, they have developed steps to employ computational methods to determine the most effective set of guides to use for targeting virus. The multiple guides can be used to target virus without the chance of binding and cutting in the human genome. These guides function to excise large sections of viral DNA, eliminating viral escape and reproduction.

How does Excision's AI/ML platform minimize off-target editing and maximize efficient on-target cutting? Excision's platform enables in silico design to analyze on- and off-target effects and integrates computational methods to design the most effective guide RNAs. In the case of off-target analysis, it was noted that Excision comes up with nominated sites and each of the chosen guides are screened including the leads for HIV without any signs of unintended

off-target editing being detected, including insertions, deletions, or any kind of recombination at these sites. Wet lab work follows to verify that there is no evidence of these unintended edits. It was also mentioned that machine learning is used for better and more accurate prediction and modeling of CRISPR cutting. The formula followed was to design assays, measure cutting, and input the data into the algorithm trained on this initial data set to determine key features that correlated with very successful, high activity and specific cutting. Then they can choose new features to input there. What is different with Excision's technology is that when looking for similar sites in the genome, the guides they develop are very orthogonal and therefore they noted they have yet to see off-target cleavage. For Excision, it is less about measuring or predicting the frequency of off-target effects but more about putting together large datasets that can be used for choice of guides to help direct these high-throughput assays they continue to use.

Excision's AI/ML drives better translation from preclinical to clinical due to highly conserved virus sequences. Excision noted that during the first step of viral database analysis looking at conservation and modeling cutting, they look at lead guides put forth and how they would work in different populations. When they look to target a virus, they look at the target patient population, their medical need, the targeting strategy and proteins in the specific regions of the sequences pulled that they want to excise and delete. They use a multifaceted approach to gene editing using input from a team of virologists, biometricians, and gene editors to address and excise the virus from a range of current indications. They also performed preclinical studies in animal models of humanized mice that can be infected with HIV and show robust viral replication. Post treatment, they looked for eradication of the viral stores and demonstrated they were able to give a functional cure. They noted that the second model they used preclinically was non-human primates to study biodistribution and safety. These guides were SIV because rhesus macaques couldn't be infected with HIV.

What are some evolving functions of the AI/ML platform that are being addressed right now? Excision notes that over the last 20 years the field of sequencing and computational means has evolved, so they continue to update their techniques both computationally and assay-wise to improve the ability to look for even rarer events in compliance with FDA regulations. They employ both publicly available and treatment center databases to provide additional sequencing to add to their ever-growing sequence database to find better methods to understand the complexity of the virus. They particularly focus on gathering information from literature and outside collaborations to continue adding features to their model to know more about targeting with multiple guides to predict the most effective way forward, and also internally using different nucleases. There are also certain aspects that are harder to model well computationally in animal models. Through increasing their data sets and refining and testing the data they can increase the ability to model effectively.

Exscientia (EXAI, not covered)

What is similar and different between Exscientia's AI/ML strategy/platform and those of its competitors, and how do these differences contribute to better molecules with higher probability of clinical success? While it's exciting to see a plethora of pharmatechs and other tech-enabled companies emerging within the space, the CEO highlighted that when Exscientia was formed (over 10 years ago) they were really the only players in this space. Throughout this time, Exscientia's driving mission has remained constant: to redefine the old-fashioned, linear, sequential way of thinking about drug discovery by fundamentally reengineering the process to more efficiently create drugs. To accomplish this, the Exscientia executives outlined three key pillars driving their differentiated approach (more below): (1) end-to-end learning; (2) patient centricity; and (3) allowing AI to drive hypothesis generation (ie, generative design). Importantly, this approach has been validated through the company's impressive collaborations with large pharmas (eg, Bristol Myers Squibb [BMY, not covered] and Sanofi [SNY, not covered]) and on the clinical front, by showing improved outcomes in clinical trials.

- **End-to-end learning:** What distinguishes Exscientia's approach is that they apply an end-to-end process with not just a single algorithm or single new technology, but rather with an entire suite of algorithms and technologies that are combined to create a patient-centric, AI-first approach from ideation to creation of a new medication and its approval.

- **Patient centricity:** A key driver for Exscientia is the patient first concept. In essence, how does the company take the needs and understanding of true human biology (eg, actual patient tissues) to a target profile for a definitive drug to drive this forward? The patient-centric approach leverages deep learning approaches and single-cell analyses of human tissues.

- **Generative design:** Given the wealth of data that is now available and the vast understanding of chemical space, Exscientia believes that AI is fundamentally better suited for making decisions regarding drug design. The concept of generative design (algorithms designing molecules and driving them forward) is central to what Exscientia has pioneered over the last 10+ years and is necessary given the cognitive bandwidth of data incorporated into drug design. In addition, AI has been instrumental at solving key complex design challenges that have gone unsolved for years, including the design of bispecific small molecules (molecules that can deliberately target two different proteins simultaneously). Moreover, by using AI to design and select molecules, Exscientia has seen far more efficient processes, with ~90% fewer molecule compounds needed for selection for testing.

What does it mean to be patient-centric? Exscientia uses a number of internal technologies, but one important proprietary tech that the executives highlighted was their automated high content patient tissue platform. Through this, they can capture live tumor samples and place them in an *ex vivo* environment where they are able to maintain the tumor microenvironment; these models are then used to test drugs during the discovery phase (similar to *ex vivo* clinical trials). This approach enables Exscientia to remove biological variables associated with the classical approach, which involves creating an immortalized cell line from a single cell. By applying single cell resolution imaging and AI to these primary tissue disease models, Exscientia is able to see how both the cancer and immune system reacts to a compound during discovery, and effectively run these data in a prospective clinical trial. Importantly, this approach was validated in a recently published study that showed improved clinical outcomes vs physician choice in late-stage cancer patients. With this validated model, Exscientia is able to understand how its early stage compounds impact heterogeneous patient populations during early discovery stages, and thus, allows them to think about patient selection/stratification and clinical trial design during the early stages of drug discovery. In addition to late stage clinical decision-making and patient stratification, these same models can be used for target discovery, where they can compare the heterogeneity of responses amongst different patient populations and identify targets that have the highest probability of response. These validated models can also be used to test hypotheses generated from the Centaur Biology platform (integrates a very large amount of external data: 30M+ abstracts in PubMed literature, publicly available genomic information, etc.), and test these hypotheses on a genome-wide scale.

Applying the right technology for the particular problem. The executives pointed out that drug discovery is plagued by both big data and small data problems. Big data issues come from the wealth of data available, and small data issues are associated with the lack of available data on novel targets. There is no one technology that the company is committed to — instead, Exscientia is focused on applying the appropriate technology for the respective problem on hand (be it small or large), and creating a system where all of these tools can be applied with the ability to loop learnings and move between different systems.

How do partner companies vet Exscientia's technology prior to collaboration? First, external companies will look at the compounds created by Exscientia and try to determine if that particular compound is able to solve problems they aren't able to address internally. Once they see these compounds, companies will then consider the technology and how it was created. The executives pointed out that for both BMY and SNY, the larger multi-billion dollar contracts came after initial smaller partnerships, suggesting that once companies better understand how the technology works, they are more interested in larger scale partnerships.

Generate Biomedicines

Generate is using AI/ML to program DNA to generate novel therapeutic proteins. Generate is combining ~200,000 high-resolution public protein structures and next-gen sequencing with

190 million amino acids worth of sequence data to recognize generalizable principles that govern protein function. Generate's machine learning platform aims to understand the complex relationship between protein sequence, structure and function. By training the platform on the collection of protein structures and sequences found in nature, supplemented with proprietary experimental data, Generate can learn the rules by which amino acid sequence encodes protein structure and function. In doing so, Generate is applying AI/ML to construct novel proteins that nature hasn't sampled.

De novo protein generation is the application of computational modeling to design a protein that binds a given target with precision and specificity to the epitope of choice. Mike Nally, CEO described, "The Holy Grail is to instantly generate optimal molecules to an existing or emerging threat. But the technology is still evolving." Generate integrates its AI/ML platform with the wet lab to dramatically increase productivity. Mike explained, "The beautiful thing about computational platforms is the scalability. Historically, companies could make around 100 protein variants. Today, Generate can make upwards of 10,000 variants and test each through cell-based assays." Importantly, the system is iterative where all of the data is fed back into the AI/ML platform, which is learning to design better protein faster on the next run. As a result, Generate noted it is decreasing the time and increasing the success rate of protein discovery.

An example of this is the company's COVID program. With the emergence of omicron variant last fall, Generate sought to generate novel antibodies to the most conserved regions of the SAR-CoV-2 spike protein, specifically Class III and Class IV regions and the S2 domain. In a 53-day learning loop, Generate was able to generate 2,500 unique full length IgG antibodies, many of which neutralized omicron and delta variants better than currently approved antibodies. In an iterative process, Generate is now conducting subsequent learning loops to further improve these constructs. This work is not only relevant for COVID, but can also be applied to other infectious diseases such as for RSV, Flu and even HIV.

Co-Optimization. The Generate platform can also be used to optimize and improve existing proteins. Beyond affinity, Generate is co-optimizing proteins for other characteristics such as immunogenicity, manufacturability, and developability. One example Mike highlighted was to improve affinity nearly 50X of a strong 60 picomolar binder to 133 femtomolar binding! This enhancement could fundamentally change the dosing (monthly to quarterly) and safety of therapeutic proteins. Another example Mike shared is asparaginase for the treatment acute lymphocytic leukemia (ALL). Asparaginase is a 200AA long protein that is highly immunogenic. Generate was able to change 70% of the sequence to reduce anti-drug antibody (ADA) formation, while retaining enzymatic activity! The opportunity is for Generate to create the most desirable therapeutics as quickly as possible.

Integrating Cryo-EM Structure and AlphaFold. Knowing how important co-crystal structure and protein-protein interaction data is, Generate is building one of the largest Cryo-EM facilities in the world with 4 cryogenic electron microscopes in Andover, MA. This facility will enable Generate to create protein structure for each variant synthesized, thereby producing enormous quantities of data to further teach the AI/ML platform. Additionally, Generate has integrated DeepMind's *AlphaFold* into the platform to help determine the folding patterns of these novel protein structures. All of these additional capabilities, iterative learning loops and integration of data have the goal of generating the optimal protein faster.

Transformative Amgen Alliance. In January 2022, Generate entered into a multi-target alliance with Amgen (AMGN, Raymond). Generate received $50 million upfront and is eligible for $370 million each for 5 programs ($1.9 billion total), with an option for an additional 5 programs for additional funds. Generate is eligible for low double-digit royalties, and Amgen will invest in Generate's next financing round. As Mike described, "This alliance marries Generate's unique discovery capability with Amgen's disease area, clinical development, manufacturing and commercial expertise." Mike went on to say that this deal enables Generate to do more. With no limit to the computational capability, Generate expects to be able to enter into additional partnerships and advance its own proprietary therapeutic pipeline.

Immuneering Corporation (IMRX, Malekar)

How does IMRX platform differ from its competitors? IMRX is applying its expertise in bioinformatics and computational biology where the platform is built on 4 main pillars including insights from human data (transcriptome analysis), novel biology (includes disease canceling tech), novel chemistry (includes Fluency), and translational planning (includes humanized models). While providing the history of the company, mgmt noted that they have been in business for over a decade with initial focus on identification of transcriptomic signals from patients that respond well to certain class of drugs. Historically, IMRX has been functional without VC money owing to the revenue raised by pharma partnerships. These partnerships were also noted to bring in deep expertise (from pharma partners) where data was thoroughly interrogated in a way that the resulting output could be used to identify viable targets/molecules. IMRX uses its proprietary disease canceling technology (DCT) to identify transcriptomic signals that drive the disease, and thus identification of novel targets. These outputs can then be fed into the Fluency platform that consists of 2 inputs: amino acid sequences of the target and a library of the molecules as potential hits. Fluency uses deep learning architecture to identify interactions between the molecules and proteins and the output is rank-ordered based on these interactions. The unique characteristic of Fluency is recognition of the specific site of interaction and its potential effect on hampering the disease signal. Lastly, the platform only needs input of SMILE (Simplified molecular-input line-entry system) strings that allows for assessment of millions of compounds in an afternoon. IMRX platform also consists of proprietary in vitro 3D tumor models that mimics the human disease and increases the probability of clinical translatability.

IMRX platform has identified shortcomings of approved MEK inhibitors and aims to address these gaps. Transcriptome analysis of approved first-gen MEK inhibitors revealed that these drugs had a potent anti-tumor effect at initial 3-12 hrs timeframe but led to amplification of disease causing signals at the 24hrs time point and raises major safety concerns. This observation led to the idea of lead molecule, IMM-1-104, that has shown deep disruption of the oncogenic signaling (in preclinical studies) with its small half-life avoiding safety concerns. The cyclic inhibition method and the unique binding of IMM-1-104 has also shown prevention of resistance mechanism associated with CRAF bypass (more details in our initiation report). It was noted that IMM-1-104 has produced a superior tolerability profile that showed only 3-4% tumor-adjusted body weight loss vs 13-17% for selumetinib and binimetinib, respectively in KRAS tumor model. Recall, IMRX is on track to file IND for IMM-1-104 in 3Q22 and enter clinic in 4Q22. We remind that IMRX has also recently designated IMM-6-415 as the lead development candidate for MEKio which is currently being tested in IND-enabling studies with an expected filing in 2023.

IMRX platform is agnostic to the target, mutation, and the disease. While speaking about the broad applicability of the platform, mgmt noted that the 2 neuro programs in Alzheimer's have used Fluency for target/hit identification (with very little changes while switching from the oncology indications). This was achieved by looking at the distinct subset of patients based on their gene expression, followed by data input and processing via DCT to identify novel targets. These targets are validated by Fluency along with hit generation with potential disease modifying effects. Mgmt believed that there are no limitations to the use of their platform in oncology, neuro, or other broader indications. Mgmt also noted that they use libraries of compounds that are pre-screened for drug like properties and after processing through Fluency, medicinal chemists assess the physicochemical properties to identify any solubility, stability, or exposure related concerns early on in drug development. Mgmt also noted that they have a patent pending for their AI/ML platform with output that aligns with the pharmaceutical quality in terms of statistical and experimental validation.

Insitro, Inc

How does Insitro's approach differ from other players in the space? Mgmt noted that AI/ML tools are broadly used and mentioned that the general approach is divided into 3 buckets of interest, including target, molecule, and patient. It was also noted that majority of work in the space is done in molecule design (small and large molecules), capsid design for AAVs, and some presence in novel target discovery. It was also noted that industry is lacking disease modifying targets and that many drugs fail in PhII/III as they do not modify the disease; one of the key focus of Insitro is to assess such disease modifying targets that could potentially reduce the number of

failures in clinic. Insitro is using human data to discover novel targets and noted that the sources they use are relevant and serve 2 purposes: (1) data from humans is interrogated to understand the effects of gene perturbation and impact on biomarker traits that reveal underlying pathology (eg, EKG, histology) and can be used as fairly simple numerical traits; (2) using genetics as underlying driver of the disease and then use iPSCs to capture these traits (initial focus on neurons in CNS indications and hepatocytes for liver diseases). It was noted that by looking at the genetic burden of the disease, one can quantify high and low burden disease and Insitro undertakes such genetic evaluations and uses to bridge human data and cellular systems that can allow identification of disease modifying effects.

What are the nuts and bolts of Insitro approach to hypothesis generation and assuring high translatability in clinic? The hypothesis generation method was noted to be driven by a combination of lab experts, insight from human data, and the ML platform itself. Wet lab comes after ML has created a model system that has identified different targets and can be introduced in cellular system (to test the hypothesis). The ML capability also helps to prioritize these targets and highlights untoward effects that may be associated with a particular target, thus allowing for a safety screen at the same time. The cellular systems are then tested to see if altering a particular target can revert the disease. Mgmt noted that that integration of entire discovery to development module in a single in silico tool is far away at this point. While speaking about clinical translation, it was noted that Insitro's approach is driven by human genetics that increases confidence that it will generate hits that are relevant in humans. Also, biomarkers identified during this process are expected to be very relevant to the disease and can allow for patient selection criteria, and in turn, an enhanced effect size of the therapy. Mgmt noted that drugs with disease modifying effects have failed in the clinic due to heterogeneous patient population in certain disease areas. For example, 15-20 yrs ago breast cancer was considered to be one disease and with several discoveries over time, now we know that they can be classified into several subtypes such as HER2+ or TNBC. Insitro plans to identify such traits early on that could allow for a robust biomarker strategy, endpoint selection, and higher probability of success in clinical trials. Such methodical approach could also allow for to see larger drug effects with smaller clinic trials that can be done at low cost and at a faster rate.

Insitro's approach is disease agnostic with initial focus on CNS and liver diseases. While talking about their approach in CNS, it was noted that one of the challenge in neuronal diseases is the thorough understanding of various mechanisms involved in functioning of a neuron (with a lot of overlap in signals adding to the complexity). It was concluded that more the amount of neuronal data you have, the ML platform will have a better predictability. Furthermore, hepatocytes are a completely different cell type that will require different set of data to build the prediction tools and that there are ways to design ML model that allows to extract insight from diverse aspects, thus providing the capability of multitask learning. Before looking into cross over effects and utilization of the same platform across different cell types, mgmt noted that it is important to look at one cell type at a time and maximize the understanding before making such switches. Mgmt also mentioned that one of the strengths of ML platform is visual data where they have invested in microscopy and developed proprietary tech to identify traits in live cells via high resolution data content. Insitro is also looking at human data (using UK BioBank), histopathology data, and partnered with Genomics England (that consists of 10k cancer patients) to strengthen their database. Finally, mgmt pointed that data from RNA-seq and Microelectrode arrays can be used to assess neural signals and further strengthening the database. To build ML algorithms for liver specific indications, Insitro has collected blood biomarker and histopatholgy data from Gilead (GILD, Kim) while publicly available brain MRI data can be used for CNS related indications. Although different cell types, mgmt noted that there can be some overlap between liver and CNS that can be leveraged by AI/ML; for eg: inflammation that is driven by over activation of immune cells (kupffer cells in liver and microglia in brain). Insitro will capture the common effects and then try to identify how it can be used to a more rapid path of identification of novel targets for different diseases.

Where does Insitro see the field 5 years from now? It was noted that ML has made tremendous progress to date and the platform is applied to a more challenging domain like

humans recently. It was also noted that there is large complexity associated with the use of this tool in humans and one needs to be very thoughtful and careful as the stakes are very high (due to implications to the human life). Mgmt was excited about 3 areas and believed that it could shape the future of AI/ML: (1) ability to transfer insight from one task to other with provisions to be able to work with limited data, (2) knowing what one don't know and quantifying uncertainties with extra caution on jumping to conclusions; (3) investigation of causality and not just correlation where adopted learnings from genetics should be taken with a grain of salt and its causal effect on disease should be thoroughly assessed before making decisions.

Metagenomi

Metagenomi's proprietary AI/ML platform uses metagenomic data to discover next-gen gene editing tools. Metagenomics is the process of characterizing the genetic material of all microbes that exist within a natural environment. Metagenomi collects samples from diverse natural environments around the world and performs next generation sequencing (NGS) on microbial DNA to generate large metagenomic libraries. This requires massive computing power. To put this in perspective, Chris Brown, Ph.D., Senior Director of Discovery, explained that "one gram of soil contains thousands of bacterial species, and even more viruses and mobile elements, creating approximately 1 terabyte of genomic data." Metagenomi collects many diverse samples; the company has collected >310 trillion base pairs, assembled into 8.6 trillion bases, equal to >2.9 million microbial genomes, predicting 3.6 billion proteins. Metagenomi then uses AI/ML computational tools to identify novel gene editing tools for therapeutic applications.

Why gene editing? While CRISPR offers broad therapeutic potential, there are limitations to current gene editing approaches, including off-target editing, large nucleases that create delivery challenges, lack of translation across disease states, and pre-existing immunity. Bacteria and archaea have evolved complex adaptive immune responses to fight viral infection. These CRISPR systems store small pieces of viral DNA or RNA, tagged as foreign to recognize and target for nuclease-mediated cleavage. Importantly, the microbes have evolved multiple different systems and nucleases targeting RNA, single-stranded DNA, and double-stranded DNA viruses. One obvious characteristic is that smaller nucleases can be more easily packaged into delivery vehicles. Further, much of the research in the field has been conducted on human pathogens, thus creating pre-existing immunity in some patients to the components. Since Metagenomi extracts samples from diverse environments, there is a lower likelihood of human exposure and thus pre-existing immunity. Metagenomi applies proprietary AI/ML algorithms to look for patterns across the data sets to find the novel tools for gene editing. Metagenomi has already identified more than 100 novel families of nucleases using metagenomic data.

The right tool for the right job. Metagenomi is creating a large "toolbox" of gene editing systems. As Chris explained, "Sometimes you want a double-stranded DNA break…But maybe you want to modify RNA or integrate a piece of DNA." Bacteria and archaea have evolved systems that have these capabilities, so that it's just a matter of identifying the right one for a given application. Metagenomi is developing next generation CRISPR nucleases, base editors, site-specific DNA integrases (CASTs), and transposases. Simon Harnest, Chief Investment Officer and SVP of Strategy, stated, "There isn't just one set of gene editing tools that will work for every disease. Metagenomi is differentiated in the fact that we can pick and choose from a variety of different nucleases, leads, and guides to tailor therapeutics to virtually any disease." For Metagenomi, it's about matching the right editing tool with the right delivery system to optimize therapeutic outcomes.

Metagenomi is applying novel gene editing tools to *in vivo* gene editing and cell therapy. Metagenomi's lead program is for Hemophilia A and comprises an AAV encoding Factor VIII gene + an LNP encapsulating nuclease mRNA and guide. Metagenomi is also developing an undisclosed LNP-encapsulated liver target KO program, potentially part of the Moderna (MRNA, Tenthoff) collaboration, as well as a research-stage cystic fibrosis program partnered with the CF Foundation. For cell therapy, Metagenomi has the capability to perform multiplexed editing at high efficiency. While Metagenomi is developing an undisclosed proprietary cell therapy program and ultimately envisions *in vivo* CAR therapies, the company will partner leading cell therapy companies.

Transformative Moderna Alliance and Future Partnerships. In November 2021, Metagenomi entered into a multi-year alliance with Moderna to develop up to 13 therapeutic programs. Metagenomi received an undisclosed upfront payment and "large" equity investment, and retains 50:50 co-development/co-commercialization rights on the lead program. Beyond Moderna, Metagenomi sees the opportunity for additional large deals with industry-leading players, as well as multiple cell therapy partnerships. Further, Metagenomi could partner for novel delivery modalities.

Odyssey Therapeutics

How does Odyssey's AI ML platform differ from its competitors? Odyssey noted that AI and ML, used with a suite of additional capabilities, are tools to address drug discovery questions. They understand the contextual value of their capabilities and that some situations will be better served with one over the other, therefore ensuring that their tools are properly used throughout the value chain. Mgmt notes that big pharma companies can have a hard time integrating their technologies when they are not well adopted throughout the organization, which contrasts to Odyssey. Mgmt shares that throughout the company, their team truly believes in the technology which is fully adopted end-to-end. The teams and workflows are constructed in a way that the chemistry, biology, preclinical and early development work in unison from the start and only use tools when appropriate, which has proved to be extremely successful. As such, Odyssey has been making significant investments for the application of AI and machine learning to computational drug discovery, believing this field will rapidly evolve in the coming years. Specifically, Odyssey is focusing on quantum machine learning and ensemble modeling as these approaches, unlike traditional ML methods, require significantly less data to train a model. This enables them to tackle many types of problems that up until now, have been less applicable due to lack of available data. Next, Odyssey keeps the human in the loop throughout all of their processes. The team collectively possess tremendous experience in drug discovery and development in that their scientists have put well over 100 molecules in the clinic and 30+ drugs on the market. By integrating computational approaches, machine learning, and the human expertise, the company can best use its resources. Because of the Odyssey team's prior experience in machine learning, quantum chemistry, and quantum physics, they believe they can effectively combine the depth and breadth of their knowledge. Their people have already done a significant amount of foundational work to understand their respective fields. In terms of the integration of biology, chemistry, and quantum machine learning, Odyssey believes that this can be applied to target selection for molecule design. Then, they seek to pick the right patient in clinical trials, which is an ambitious and long-term goal. Currently, they are focusing on molecule designs despite having future plans to move their platform to clinical trials and ultimately target selection. Even for programs that did not come out of AI and ML, these tools may still be implemented. When they are interested in a particular target that possess some literature data, they can take this noisy data, train a model and then use it to interrogate a large virtual, commercial compound library. In a specific example that took about 6 weeks, mgmt recounts that approximately half (out of the ~100 compounds purchased) were active, many of which showed single digit nanomolar potencies and several orders of magnitude selectivity over close anti targets. Importantly, they are able to identify molecules that had no prior biological activity, which mgmt believes addresses a criticism of AI/ML, ie, this type of exercise only finds things that have already been found and close to the data set.

What is different with Odyssey's DEL library. Odyssey has 9 internal programs in the pipeline and hopes to enter the clinic with several of them in the next couple of years. Their capabilities support their entire portfolio of biophysics, structural biology (proteomics), AI and machine learning. They also have a suite of technologies dedicated to their specific products/projects. Within their small molecules portfolio, they are focused on the ability to rapidly generate hits to identify chemical matter for many other programs. Most notably, Odyssey has made a significant investment in developing proprietary DEL libraries, which helps them generate novel solutions that they would not have been able to do otherwise. These libraries are particularly in areas of covalent DELs, which provide covalent chemistry for cysteine, lysine, tyrosine, and other crucial amino acids and molecular proximity, especially regarding stabilization of protein-protein interactions. Odyssey builds these libraries in an iterative fashion, using the computational

approach in tandem with their human experience. Next, Odyssey improves these libraries with ML that identifies trends and parameters. These are then employed with unbiased generative modeling capabilities to enhance library design, which is unique to Odyssey and effectively facilitates high quality libraries that find better solutions, faster. The ability to sample a more diverse chemical space will thus open more solutions for a wide range of problems. Accordingly, mgmt is excited about this natural products approach, because for the first time, there is a high throughput screening capability to look at microorganisms and their functional, natural products, in a target specific manner.

Power of harnessing AI/ML in identifying compounds targeting covalent inhibition. First, Odyssey reminded us of their chemistry focus and of their strong team of chemists, before sharing the importance of covalent chemistry. For targets that only have allosteric pockets or shallow binding sites, covalent chemistry is an effective way to drug those pockets in the absence of orthosteric binding sites. In addition, covalent chemistry often can decouple pharmacokinetics and pharmacodynamics. This means that a smaller dose can provide long-term benefit without the need for target coverage and pKA, as would be otherwise necessary. Odyssey notes that covalent chemistry has offered an attractive target product profile because they have the in-house expertise to build chemistries for pockets with amino acids, beyond cysteine. In some cases, they have found that among their different types of libraries, their investigated targets cannot be addressed with anything besides a covalent approach. Next, they couple ML to library designs which allows them to enhance their libraries in ways that are not possible with chemical intuition alone. This integration provides them with a unique insight into chemical matter. Odyssey believes the chemistries they use in-house to construct libraries differentiate them from others, in terms of their molecular diversity and physical properties.

Odyssey sets a new standard — DME with 60% improvement; here is how they do it. Mgmt notes that this is possible through quantum ML and the ensemble scoring functions that they use. Compared to a traditional ML that usually applies two-dimensional data (SMILES strings or things not rich in stereochemistry), quantum mechanical ML uses three-dimensional qualities that provide a molecule's quantum mechanical features. By combining this level of three-dimensionality with their hybrid scoring function, this includes multiple endpoints all built together. Furthermore, quantum ML and ensembling handles noisy data extremely well and allows them to use smaller training data sets on the orders of 10s-100 (compared to 1,000s or 10,000s). Together, these features allow for better solutions and Odyssey then continues to drive those forward.

Recursion Pharmaceuticals (RXRX, not covered)

What differentiates Recursion's AI ML platform from its competitors and what makes these differences critical to success? RXRX management noted that there are three primary differentiating factors, and in order of importance: **1) Integration of wet lab with dry lab at every step of the process**. This allows them to run actual biological experiments, which facilitates an iterative cycle of innovation and experimentation at a massive scale. **2) Recursion is a "biology first" company.** They emphasize the relationships between potential targets and the rest of the body, as an entire system, rather than only considering the individual chemistries. **3) Scale of their proprietary dataset.** As the size of the dataset itself grows with more experimentation, so do network effects, because the program makes relationship predictions of biology and chemistry. Now, across 100M+ experiments, the power of scale is evident. While there are other companies that have each of these characteristics, Recursion differentiates itself by possessing all three. More specifically, Recursion's digital high dimensional biological data systems are unique due the scale of omics data that are all done in-house; it is bigger than the publicly-available datasets and is completely proprietary. In addition, RXRX outlined that imaging cell morphology via computer vision and ML is the foundation from which sophisticated signatures of morphological features like shape, size, texture, and spatial distribution of organelles can be extracted. To the best of their knowledge, no other company has a comparable large omics dataset. In fact, a pre-competitive collaboration among some of the largest pharma companies plans to build an amount of cellular morphology data over two years that only takes Recursion 2-3 weeks to generate. Furthermore, simply building a genetic map is not enough for other companies to overcome their proprietary data sets because there is more to it. Recursion has

built sophisticated computational algorithms that pair with large proprietary data sets, so even as ML and AI enter this space and data becomes commoditized, it will be challenging to use without robust complementary tools. Only the companies who are meshing sophisticated computational approaches, like convolutional neural nets with well-controlled tools specifically made for ML data, will be the most successful. Few others have built such a system and of those companies none that they are aware of have the same scale as Recursion, offering them an advantage and some protection as AI and ML is further adopted.

What is the end-to-end process of integrating computation and wet lab for RXRX? The approach is cyclical in nature as the results from each side feeds into each other. Each week, they run up to 2.2M experiments to expand their map. This typically entails adding proteins to primary human cells (PBMCs, specifically), that are then co-cultured to quickly establish complex cellular systems. To demonstrate a starting point, Recursion has a complete, proprietary CRISPR library from designed experiments where every gene in the human genome has been knocked out, in multiple cell types. The images from omics data are fed into the computational side of the company, where neural networks that are trained from prior data are used to interpret and elucidate relationships across biology. Accordingly, they filter through datasets to formulate hypotheses before testing in vivo. Filtering begins with an incontrovertible piece of biology as an anchor that places an assumption. In the specific example mgmt gave, following literature review as a starting point, their scientists identified CDK-12 as an interesting target for homologous recombination-deficient (HRD) ovarian cancer. Querying the map to find related targets involved in a known piece of biology across the entire genome reveals genes or molecules might not be known to associate with the investigational target. In the case of CDK-12, they discovered a newly-associated gene, which prompted them to further filter this result to see other profiled compounds that may affect the novel target. To close this experimental loop, they verify hypotheses in the wet lab, using biological models (such as organoids, spheroid, patient-derived cells, and animal models). More on the CDK-12 example, they initiated an animal study with an HRD-negative ovarian cancer line and following the readout, found a 100% complete response. Because Recursion has statistics in their S1, the ML predictions are accurate in the animal model ~30-50% of the time, which is important considering the cost to build these maps. Furthermore, failure is acceptable when it happens early and few resources are invested. Contrasted to failure in a typical approach that demands significant investment prior to realizing there is no potential, Recursion is building a system to obtain prompt verification early in the process.

When extracting external data, what quality controls are in place to maintain the map's accuracy and what further conclusions can be driven? Recursion noted that most academic literature is not completely accurate and its conclusions often cannot be recapitulated, which poses significant challenges. As such, they have built the data internally to avoid this issue. In relation to a novel target, Recursion use the literature merely as a starting point, and follows up with sufficiency experiments in cell cultures and animal models to elucidate if their hypothesis is valid. Next, mgmt pointed out that in the pharma industry, when a potential new target is first identified in a major scientific journal, several companies race to validate it and subsequently advance molecules, if correct. Recursion, on the other hand, is interested in investigating the unexplored and has the power to do so with the data maps because it quickly exposes relationships in biology and continues to improve these as it receives more data. Mgmt told us that their cerebral cavernous malformation study, which began PhII just two weeks ago, presents a considerable opportunity from a rare disease/unmet needs perspective. While it is six times more prevalent than cystic fibrosis in the US, there is no medical treatment for the disease. While no published study has identified their target as having potential, no commercial entities with sponsored trials are in the clinic, but Recursion used its map to find it.

Schrodinger, Inc. (SDGR, Kim)

Physics-based methods differentiates platform from AI. Schrodinger's physics-based software platform relies on first principles or established laws of physics, as opposed to empirical or statistical modeling and fitting parameters of AI. Physics-based methods can predict properties of novel molecules for which there is no previous knowledge, thus without the need of a training set. The ability to extrapolate into new chemical space sets the technology apart from machine learning, which as a knowledge-based method can only make predictions about what is already

known and similar to the training set. Schrodinger emphasizes the competitive advantage in tailoring characteristics of a molecule relative to the ideal target product profile, citing its discovery of an ACC inhibitor as an example of a design challenge that the platform had overcome. Physics-based methods can predict multiple product properties (eg, potency, solubility, permeability), which are complex and sometimes anti-correlated to one another, enabling optimized drug discovery with a higher probability of success.

On the path to overcoming limitation in structural biology. Physics-based methods require correct and accurate 3D structure of a protein as a starting point for drug discovery. However, the number of high-quality protein structures is growing in recent years and expected to increase exponentially due to advances in technology, including cryo-EM, X-ray crystallography, and computational methods. Schrodinger has also acquired XTAL Biostructures, a structural biology company, to accelerate internal initiatives to identify target structures that are highly compelling. While physics-based calculations can be computationally expensive and time-consuming, Schrodinger also leverages the accuracy and extrapolation of physic-based methods with the speed of machine learning. Management believes AI will not replicate or replace physics-based methods, because of the infinite number of ways to combine organic elements, but for AI to improve, physics-based methods and machine learning must work in tandem as more data is generated.

Internal pipeline will be therapeutic area-agnostic with multiple modalities. Lead internal programs are oncology-focused, with MALT1 near clinical start and CDC7/WEE1 not far behind. Additional oncology targets remain appealing, with kinase programs, transcription programs, PPI, and resistance opportunities as areas of interest. Expansion beyond oncology for future development includes immunology, with management specifically noting genetically-driven immunology and modality-switching as good focus areas for the platform, while more complex therapeutic areas such as CNS are regarded as opportunities for partnerships.

Shape Therapeutics

How does Shape's AI/ML platform differ from its competitors? Mgmt highlighted that the platform is intended to develop RNA-based therapies and reminded that they had started looking at it 5 years ago, and it has been recently validated with the COVID vaccines. The concept of Shape was to merge AI with RNA, and understand the footprint of RNA and how it works, leading to the evolution of a new field called "programmable arguments." Shape is building RNA-based therapy across multiple modalities to be developed internally and deploy some to the rest of industry, thus allowing a broader patient population to be addressed. Mgmt noted that they are collecting tremendous amounts of data, not only in RNA editing, but also in payload delivery, which is driving the decision to manufacture payloads simultaneously in-house. Shape is leveraging AI to understand the rules about what makes a good target, what makes a good delivery system, and then use that information to retrain experiments and subsequent refine them to improve the predictability. Furthermore, Shape differs from others where they are turning biology into a physics problem and reducing it to the first principle; in the case of RNA-based technology they are applying these principles to attempt to cure Parkinson's or Alzheimer's. It was also noted that for such diseases, therapies may comprise 1,200 nucleotide RNA (bigger than the number of atoms in the universe) and their AI/ML platform can the capability to support this, backed by a massive library. Compared to others, Shape's AI platform uses different deep architectures and unique in-house technologies, supported by strong team of bioinformaticians that can account for nuances in next generation sequencing (NGS) data. Converse to classic biology, Shape identifies a "winner" (generated by the platform) based on several patterns that can help infer new rules that start explaining if RNA works.

Shape's AI/ML platform is modality-agnostic and has identified 3 RNA therapies so far. One of the modalities is to edit a gene mutation, the second is to target a specific type of stop codon using tRNA, and the third is replacement of RNAs. CRISPR was noted to have limitations in certain cells (like liver and blood) and editing tools using foreign protein can illicit an immune reaction. Targeting RNA does not have this liability and does not require protein delivery, and Shape is leveraging ADAR (naturally occurring) which allows for reallocating the ADAR to make an edit elsewhere in the transcriptome and has applicability in non-dividing cells (like neurons

and muscles). Shape's platform is equipped with RNA skip, RNA fix, or RNA swap technology, allowing it to use the best technology and delivery system for a specific indication. ADAR doesn't recognize a sequence but instead recognizes tertiary structure of RNA, and the effect of its transcriptome editing cannot be done in traditional ways that requires AI/ML capability to find sequence patterns. Shape runs a screen in-house with 100s of 1,000s or a million guides, and can identify top winners that achieve 90-95% on target editing with some off-target editing (edits nearby adenosines). Furthermore, Shape uses its deep learning architecture to analyze that data and then generate a novel guide with a new RNA structure, and achieves nearly 98% on target editing, with zero off target editing of nearby adenosines. This is critical, because for this technology to work, there needs to be minimal off target editing. Shape uses internal de novo data and a generative model to predict new sequences that can be leveraged for novel therapies. Mgmt noted that whole world was working on one structure derived from a gene called GluR-2, but ADAR is extremely conserved (fruit fly almost has the same ADAR as human) that allowed pulling every related structure in their first library and with each evolution/learning of the algorithm, Shape has been evolving this library to be closer and closer to the simplest solution to any problem.

How does Shape's library/AAVs differ from competitors? Shape is focused on RNA therapies and their ADAR platform is just picking up steam. It was noted that every other player in the space is chasing ASOs, which may show efficacy but can have safety drawbacks. Compared to every other ADAR company, Shape believes that it is chasing something slightly different, which is the potential to provide one-time treatment. This will be accomplished by use of viruses like AAV to deliver a piece of DNA that sit in a cell forever (as long as the cells remain alive), with potential applicability in differentiating cells (ie, 95% of the body). While highlighting the shortcomings of current AAVs and LNPs (liver tropic), mgmt noted that they have identified fractions of viruses that "stick" to a particular tissue (like liver), and replacement of such components (via screening of 100 billion mutant viruses), they can achieve specific tissue tropism. This is done by injecting at a scale of a billion AAVs in NHPs and assessing the tissue interactions (done with AAV5 and AAV9), where several similar cycles can allow for meaningful conclusions. Specifically, they look at the biophysical parameters of each amino acid and capsid sequence in 48 different tissues, which allows prediction for a given signature in a specific tissue, like brain. Out of 1 billion capsids screened in NHPs, Shape has narrowed down to 50K capsids that can be utilized for brain delivery, with the goal to identify the #1 candidate.

Totus Medicines

Big data sets becoming more reliable. Totus is focused on bridging the gap between AI and ML by incorporating novel high-throughput techniques and coupling ML capabilities with proprietary techniques (eg, high-throughput confocal microscopy and DNA encoded libraries) to create larger data sets with improved quality. Additionally, Totus plans on utilizing these novel techniques to pursue a modality and drug agnostic approach. Moreover, Totus' platform could improve the efficiency of newer tools such as chemoproteomics and high throughput confocal microscopy, so chemical spaces across drug targets can be explored in a rapid manner. Totus employs ML in developing and scaling these technologies across large libraries of compounds while maintaining the reliability of the data.

The quality of the data is the key to building better programs. Older drug discovery techniques to screen molecules (eg, combinatorial and parallel chemistry techniques) have generated data that can lack in reliability. According to the company, newer techniques in the field to screen different molecules and therapies are driving higher quality and more reliable data to power and improve machine learning. The development of more sensitive biosensing approaches and analyses have led to more predictive and improved understanding of how molecules behave with targets in a cellular setting, enabling more efficient drug discovery for hard-to-treat or untreatable diseases.

Looking to advance a differentiated PI3KA molecule. Safe, targeted covalent molecules have historically been challenging to develop, but as techniques have improved, viable therapies such as Imbruvica (BTKi), Lumakras (KRAS G12C), and Tagrisso (EGFR) have been approved. The

company expects a Phase I to initiate before year-end 2022 to confirm the efficacy shown in the preclinical studies.

XtalPi, Inc

XtalPi's approach combines the creativity and speed of ML with the accuracy of physics. When we asked William Glauser (the SVP of BD at XtalPi) about what differentiates XtalPi's AI/ML platform vs other players in the field, he noted that XtalPi is the only company that combines quantum physics with AI and wet-lab automation. XtalPi is able to collect information regarding quantum physics (eg, the rules that govern the activity of molecules) by using a nanoscope with atomic resolution to understand the precise activity of a protein target and interacting compound. Through this approach the company is able to generate de novo data that can "kick-start" early research programs and be used to train ML models. Glauser explained that they have closed the loop between dry lab predictions and wet lab testing by heavily focusing on filtering out compounds that are not likely to become developmental leads. By doing this, XtalPi is able to hone in on compounds that are most likely to be developable and have a good chance of clinical success.

Widespread utility of using universally applicable physics to train AI. XtalPi is able to apply its quantum physics approach to a large range of compounds to calculate a myriad of properties, given that physics is largely agnostic to variables such as MoA or targets. Glauser noted that these approaches are applicable to virtually any compound and molecule type where a 3-dimensional structure of a protein target or anti targets (for pseudo-activity) are known.

Combining dry lab computational predictions and wet lab automation. According to the company, the platform, which utilizes high-throughput experiments with robotic automation, helps to energize and amplify human intuition by allowing scientists to focus on more difficult synthetic pathways that robots can't handle, while offloading more routine pathways to the robots. This helps to avoid the bottleneck of synthesis in drug development. XtalPi has two core pillars to its business, covering the entire chemistry value chain from discovery to manufacturing a chemical compound — the first being solid state (ie, going from drug substance to drug product) and the other being drug discovery. For solid state, XtalPi hypothesizes the landscape of solid structures in terms of stability, bioavailability, and developability, which companies such as Pfizer (PFE, not covered) use for making early formulary decisions. This is accomplished by computing X-ray structures of the crystals and computationally predicting the results — of which Glauser pointed to a 98% success rate.

Mitigating safety risks using computational modeling. Glauser noted that in any given cell, there are ~10K off-targets that could potentially interact with a compound in question. Utilizing its quantum AI/ML platform, XtalPi is able to front-load the risk by modeling the potential for off-target toxicity. As opposed to other companies in the AI drug discovery field which work from end-to-end, XtalPi believes that the true value inflection is within the chemistry, and thus, they work with world-class partners on the front-end who have "exquisite" knowledge of biology and physical genomics and are able to determine the target biology in advance.

Translation of data represents "the next big step in the evolution of AI." According to Glauser, the biggest problem today is the translation of data from cells to animals and then into humans. For developers to improve clinical predictability, Glauser highlighted three key components: (1) initially choosing the right target or targets, based on the systems biology and the downstream pathways of interacting with a particular target; (2) getting as close to the target profile as possible (an area that XtalPi is working diligently to address); (3) and coming up with better animal models that more accurately recapitulate the human disease.

Adaptive learning with harmonized, closed loop cycles. XtalPi trains their AI model with physics, where with each cycle of the drug discovery process, the predicted vs observed results are matched to improve the computational predictability of the system going forward. However, Glauser noted that this only works if the data is harmonized (eg, within a closed loop cycle). In addition to the AI getting smarter through iterations, other aspects of the process (eg, people, the process, and workflow) also improve through multiple iterations.

IMPORTANT RESEARCH DISCLOSURES

Notes: The boxes on the Rating and Price Target History chart above indicate the date of the fundamental Equity Research Note, the rating and the price target. Each box represents a date on which an analyst made a change to a rating or price target, except for the first box, which may only represent the first Note written during the past three years.
Legend:
I: Initiating Coverage
R: Resuming Coverage
T: Transferring Coverage
D: Discontinuing Coverage
S: Suspending Coverage
OW: Overweight
N: Neutral
UW: Underweight
NA: Not Available
UR: Under Review

Distribution of Ratings/IB Services
Piper Sandler

			IB Serv./Past 12 Mos.	
Rating	**Count**	**Percent**	**Count**	**Percent**
BUY [OW]	**677**	**67.43**	**289**	**42.69**
HOLD [N]	**316**	**31.47**	**63**	**19.94**
SELL [UW]	**11**	**1.10**	**0**	**0.00**

Note: Distribution of Ratings/IB Services shows the number of companies currently covered by fundamental equity research in each rating category from which Piper Sandler and its affiliates received compensation for investment banking services within the past 12 months. FINRA rules require disclosure of which ratings most closely correspond with "buy," "hold," and "sell" recommendations. Piper Sandler ratings are not the equivalent of buy, hold or sell, but instead represent recommended relative weightings. Nevertheless, Overweight corresponds most closely with buy, Neutral with hold and Underweight with sell. See Stock Rating definitions below.

Analyst Certification

The analysts Yasmeen Rahimi, Ph.D., Edward A. Tenthoff, Do Kim, Christopher J. Raymond and Swapnil Malekar, Ph.D., primarily responsible for the preparation of this research report, attest to the following:

The views expressed in this report accurately reflect our personal views about the subject company and the subject security. In addition, no part of our compensation was, is, or will be directly or indirectly related to the specific recommendations or views contained in this report.

Piper Sandler research analysts receive compensation that is based, in part, on overall firm revenues, which include investment banking revenues.

Time of dissemination: 5 April 2022 03:30EDT.

Research Disclosures

Affiliate disclosures: Piper Sandler is the trade name and registered trademark under which the corporate and investment banking products and services of Piper Sandler Companies and its subsidiaries Piper Sandler & Co. and Piper Sandler Ltd. are marketed. This report has been prepared by Piper Sandler & Co. and/or its affiliate Piper Sandler Ltd. Piper Sandler & Co. is regulated by FINRA, NYSE and the United States Securities and Exchange Commission, and its headquarters are located at 800 Nicollet Mall, Minneapolis, MN 55402. Piper Sandler Ltd. is authorised and regulated by the Financial Conduct Authority, and is located at 1st Floor, New Fetter Place East, 8-10 New Fetter Lane, London EC4A 1AZ. Disclosures in this section and in the Other Important Information section referencing Piper Sandler include all affiliated entities unless otherwise specified.

Rating Definitions

Stock Ratings: Piper Sandler fundamental research ratings are indicators of expected total return (price appreciation plus dividend) within the next 12 months. At times analysts may specify a different investment horizon or may include additional investment time horizons for specific stocks. Stock performance is measured relative to the group of stocks covered by each analyst. Lists of the stocks covered by each are available at www.pipersandler.com/researchdisclosures. Stock ratings and/or stock coverage may be suspended from time to time in the event that there is no active analyst opinion or analyst coverage, but the opinion or coverage is expected to resume. Research reports and ratings should not be relied upon as individual investment advice. As always, an investor's decision to buy or sell a security must depend on individual circumstances, including existing holdings, time horizons and risk tolerance. Piper Sandler sales and trading personnel may provide written or oral commentary, trade ideas, or other information about a particular stock to clients or internal trading desks reflecting different opinions than those expressed by the research analyst. In addition, Piper Sandler offers technical research products that are based on different methodologies, may contradict the opinions contained in fundamental research reports, and could impact the price of the subject security. Recommendations based on technical analysis are intended for the professional trader, while fundamental opinions are typically suited for the longer-term institutional investor.

Overweight (OW): Anticipated to outperform relative to the median of the group of stocks covered by the analyst.

Neutral (N): Anticipated to perform in line relative to the median of the group of stocks covered by the analyst.

Underweight (UW): Anticipated to underperform relative to the median of the group of stocks covered by the analyst.

Other Important Information

The material regarding the subject company is based on data obtained from sources we deem to be reliable; it is not guaranteed as to accuracy and does not purport to be complete. This report is solely for informational purposes and is not intended to be used as the primary basis of investment decisions. Piper Sandler has not assessed the suitability of the subject company for any person. Because of individual client requirements, it is not, and it should not be construed as, advice designed to meet the particular investment needs of any investor. This report is not an offer or the solicitation of an offer to sell or buy any security. Unless otherwise noted, the price of a security mentioned in this report is the most recently available closing market price. Piper Sandler does not maintain a predetermined schedule for publication of research and will not necessarily update this report. Piper Sandler policy generally prohibits research analysts from sending draft research reports to subject companies; however, it should be presumed that the fundamental equity analyst(s) who authored this report has had discussions with the subject company to ensure factual accuracy prior to publication, and has had assistance from the company in conducting diligence, including visits to company sites and meetings with company management and other representatives.Notice to customers: This material is not directed to, or intended for distribution to or use by, any person or entity if Piper Sandler is prohibited or restricted by any legislation or regulation in any jurisdiction from making it available to such person or entity. Customers in any of the jurisdictions where Piper Sandler and its affiliates do business who wish to effect a transaction in the securities discussed in this report should contact their local Piper Sandler representative, or as otherwise noted below. Canada: Piper Sandler & Co. is not registered as a dealer in Canada and relies on the "international dealer exemption" set forth in National Instrument 31-103 – Registration Requirements, Exemptions and Ongoing Registrant Obligations of the Canadian Securities Administrators. This research report has not been prepared in accordance with the disclosure requirements of Dealer Member Rule 3400 – Research Restrictions and Disclosure Requirements of the Investment Industry Regulatory Organization of Canada. Europe: This material is for the use of intended recipients only and only for distribution to professional and institutional investors, i.e. persons who are authorised persons or exempted persons within the meaning of the Financial Services and Markets Act 2000 of the United Kingdom, or persons who have been categorised by Piper Sandler Ltd. as professional clients under the rules of the Financial Conduct Authority. United States: This report is distributed in the United States by Piper Sandler & Co., member SIPC, FINRA and NYSE, Inc., which accepts responsibility for its contents. The securities described in this report may not have been registered under the U.S. Securities Act of 1933 and, in such case, may not be offered or sold in the United States or to U.S. persons unless they have been so registered, or an exemption from the registration requirements is available.This report is produced for the use of Piper Sandler customers and may not be reproduced, re-distributed or passed to any other person or published in whole or in part for any purpose without the prior consent of Piper Sandler & Co. Additional information is available upon request.